

Settling the AIDS virus dispute

Attempts to deal with what may be the most serious epidemic so far have been marred by two parallel disputes between the United States and France. May they now end.

Is the document appearing on page 435 of this issue the end of the four-year-long dispute between the French and US research groups working towards the characterization of the virus responsible for AIDS (acquired immune deficiency syndrome)? Or will it, when subjected to the examination its authors must expect, help to keep alive this corrosive argument? The circumstances are exceptional, as are the means by which those principally concerned hope they will now be able to live their lives and carry out their work in peace and quiet. But it may be too soon for them to be assured of that luxury, which is why there is much to be said for an enumeration of what has been achieved.

AIDS was recognized as a distinct pathological condition merely six years ago. From the outset, there was circumstantial evidence that it is caused by an infectious agent, but the proof (and practical utility) that such an agent is a virus must rest on a demonstration that the supposed agent is antigenically distinctive, that it can be cultivated in cells of some kind from a susceptible host and that it is, indeed, capable of being transmitted from one infected cell to another. (The classical need to demonstrate that a virus can infect whole susceptible animals is obviously inapplicable when the infection is almost always fatal, but a great many chimpanzees have now been sacrificed in that cause.) By any standards, it is remarkable, and a mark of the power and perceptiveness of laboratory science, that all this had been accomplished within two years of the recognition that AIDS is a distinct disease. It is even more remarkable that, within a year, the entire genome of four distinct strains of the AIDS virus (now called HIV) had been identified by the sequence in which their component nucleic acids are strung together. That so much had been accomplished so quickly is a cause for wonder. No disputation can in the long run detract from that remarkable achievement. In the long run, many now-infected people will benefit enormously from the speed with which the virus has been isolated and understood.

Front line

Two independent disputes have marred this success story. Those concerned are principally Dr Robert Gallo of the US National Cancer Institute, one the US National Institutes of Health under the umbrella of the US Public Health Services, and Professor Luc Montagnier of the Pasteur Institute in Paris. The more serious difference between Gallo and Montagnier has turned on the question of who was first to recognize that the AIDS virus is indeed a virus distinct from those known previously. The other dispute has been commercial; which of these, and which parent institution, is entitled respectively to the credit for and the profits from the use of HIV antibodies in diagnostic tests for AIDS infection? For the past four years, the second dispute has coloured (and sometimes has overlain) the first. The good news is that this has now been settled, apparently on terms acceptable to both sides. It is a splendidly imaginative notion that a third of the proceeds from the manufacture and sale of diagnostic reagents should accrue to a foundation devoted to the battle against AIDS in Africa, where the front line is drawn. The bad news is that there is now nothing left except the embers of the first dispute, over which an army of ghouls is eager to rake.

The circumstances that matter are the following. Montagnier arrived in the United States in September 1983 carrying a sample of what he believed to be an authentic novel viral isolate, which was stored in Gallo's domestic refrigerator. A few days later, Montagnier, talking at a Cold Spring Harbor symposium with Gallo as the chairman, described what he believed to be a proof that the infectious agent of AIDS is a virus distinct from the lymphotropic viruses previously described by Gallo, particularly from that then known as HTLV-I (adult T-cell leukaemia virus endemic in southern Japan), Gallo's candidate infectious agent. It seems common ground between Gallo and Montagnier that the former was unreasonably dismissive of the latter at the end of that talk; far from being applauded for his perceptiveness, Montagnier was left with the impression that he had been scorned. He nevertheless provided a second sample of his viral isolate before the year was out; his parent institution applied for a US patent on his behalf in December of that year. Gallo's proxy patent application was lodged in April the following year, attended by publicity orchestrated by the then Secretary of Health, Education and Welfare, Mrs Margaret Heckler.

Nature's knowledge of these circumstances derives from interviews conducted by members of its staff (never fewer than two) with Montagnier last August and with Gallo (and many of his associates) in September. Both men were scrupulously open, even to the point of indiscretion. There seems no doubt that Gallo's treatment of Montagnier (at the Cold Spring Harbor meeting, and his casual request for a second sample of the viral isolate on the grounds that he had failed to make the first sample grow) was at least discourteous. Much of the separate legal dispute has centred around the question whether Gallo's laboratory had made improper use of Montagnier's viral isolate to develop a commercial product. Notebook pages in Gallo's laboratory from that period are paradoxically headed "LAV" (Montagnier's first name for his virus), but there can be no telling rebuttal of Gallo's explanation that, at the time, there was no other name.

On the more enduring question of whether Gallo has taken some part of the credit due to Montagnier, there are three important things to say. First, of necessity, there is unlikely ever to be proof either way. Second, Montagnier has made an important point; it is possible to achieve great things with poor resources — and then to be dealt with dismissively; at the very least, he has struck a blow for good manners in science that people in competitive laboratories should take to heart. Third, Gallo's contribution to this enterprise has been every bit as important as the subscribers to last week's letter of support in the correspondence pages of *Nature* on his behalf proclaim: energy (and the resources of the National Cancer Institute) have been imaginatively harnessed to a good cause in a manner that demonstrates how much can be accomplished how quickly. It may not be too much to hope that one consequence will be a prophylactic for the worst epidemic yet.

That is why there is an urgent need that these hatchets, whether real or imagined (not an empty interpolation) should be buried. HIV may be characterized, but treatment of and defence against the disease lie a long way off. The scientific community



has a heavy responsibility: there is a sense in which taxpayers' generous investment over several decades in the uninhibited understanding of how things function has been directed towards the handling of emergencies such as these. Naturally, if there remain serious issues in dispute, people should be free to say what they like. Otherwise, they should get on with what may be a long job. □

Australia's basic science

The Hawke government may be more interested in an early election than in reorganizing research.

THE close parallel between the Australian and the British university systems is not in the least surprising. For much of their history, both systems have been run by like-minded people with similar objectives. Over the years, and out of once-filial regard, Australian planners have deliberately set out to adapt British institutions to their own needs. This, in particular, is why funds from the federal government are now channelled to individual universities through what is called the Commonwealth Tertiary Education Commission, designed to act as an intermediary between the federal government and the universities much as the British University Grants Committee is intended as a buffer between central government and its pensioner institutions. During the 1970s, both systems also grew quickly, the Australian more quickly than the British for two reasons — the interest of the various states in their own improvement and the general recognition of the social value of relatively open access to higher education. But the ingredient of the British system not yet simulated in Australia is that for supporting university research. There is a scheme for providing academics with research grants from an annual budget of less than A\$30 million, but the theory is that the bulk of the financial support for research is part of the recurrent university budget not in principle distinguishable from other costs, academic salaries for example.

The shake-up now recommended by the Australian Science and Technology Council (ASTEC) in the pattern of higher education support is that there should be a stronger research council than hitherto, with more money to spend on the direct support of university research projects and with enough autonomy to do so in a discriminatory fashion (see p. 431). On the face of things, this is another step in the path historically followed by the British system, but there is more to the proposed shift than that. First, a large part of ASTEC's argument is that the Australian system has now matured to the point at which untied graduate education is, or should be, a substantial part of its work. Hitherto, in the Australian universities, much of that burden has been carried by a small number of university departments, by the Commonwealth Scientific and Industrial Research Organisation (CSIRO) and by the Australian National University (unlike other universities, supported by government). But ASTEC also argues that, under the present arrangement, the uniform distribution of funds for research among universities and academics is not a good recipe for supporting outstanding projects and may be wasteful.

The strength of ASTEC's case for change is, unfortunately, no guarantee that its recommendations will be followed. The most obvious difficulty is that academics will protest (as their unions have done already). The process of concentrating funds on the projects most likely to make a mark, and on the people who conduct them, is never popular with those left on the sidelines. A more serious difficulty in Australia now is that the federal government is preoccupied with other things, perhaps an election. To embark on university reform at such a time is distracting; there is also a chance that in ways that cannot easily be foretold, it will give hostages to fortune. So the government is more likely to consult widely than to act quickly. That is why it should not forget that the reform that ASTEC is recommending cannot be left on the sidelines indefinitely. Sooner or later, the

Australian university system must follow others down the road of selectivity and concentration.

There is a curious irony in this development. Although, historically, the Australian system is a few steps behind the British, it is at least the first of them to have to decide on a recommendation that funds should be transferred from general support of higher education to the research which is a necessary ingredient of it. In Britain, the research councils have for some years been looking enviously at the third of the UGC budget nominally used for research support, but the only committee formally to have considered a transfer of funds (the Merrison committee which reported to the Advisory Board for the Research Councils in 1982) rejected the idea. But the way the present British government is moving, there is every chance that the proposal will be up for decision again quite soon; the government is soon to produce a policy paper on the machinery for the administration of higher education in which this proposal must be an essential part. It will be interesting to see whether, on this occasion, Australia is the first in the field. □

Across Japan's bows

The promised trade war has begun, but quietly. It is important that it should not begin in earnest.

THE strange agreement on silicon chips, concluded last July between Japan and the United States, must be seen for what it is — an agreement between two self-interested parties not to cause each other unmanageable annoyance. That the agreement has so quickly broken down is predictable. The weakness of the agreement was plain from the beginning — in capital-intensive industries, the 'fair market price' for a commodity can vary with exquisite sensitivity with the efficiency with which one partner's manufacturers make use of capital and labour. On the empirical evidence of the past few weeks (as of the past few years), Japan has had the better of this test of prowess. Part of the explanation of that is that the savings ratio in Japan is very large (more than 20 per cent of the average annual income is saved against rainy days); another part is that the government of Japan has worked out ways of borrowing at low rates of interest. The result is that there is no common standard of what should be the proper return on a unit of capital investment.

Simple arithmetic should therefore disarm the intending adversaries, but that is unlikely to be what happens. The United States could reasonably argue, in its present fix, that it will be better to sell something to Japan than nothing. If, several years from now, it could be demonstrated by a team of lawyers that Japan had organized non-tariff restraints on trade in 1987, the same team could probably go on to show that no substantial difference had been brought about. Even if it were the case that Japan is a dishonest trading-partner, which is probably not true, it would still be better for US taxpayers that the United States should stay in the international trading business than that it should withdraw. Yet sentiment goes the other way.

The most remarkable demonstration of that is the readiness of the British government to threaten all kinds of reprisals against Japan. Much cant is involved. The running issue is whether a British company called Cable and Wireless should be given a right to bid for a share of Japan's external telecommunications market. Curiously, the British have taken some trouble to ensure that strategically important companies do not fall into the hands of owners overseas; British Aerospace, which does a great deal of defence business with the British government, would not be allowed to fall into foreign ownership because of the requirement that the purchase of the last "golden share" should be at the discretion of the government. Why should Japan's telecommunications be so liberalized? Because, everybody believes, that is the way the wind is blowing. So why not let the same wind blow more freely where it might have a greater influence, in the United States? Meanwhile, why not call a truce on chips? □

US initiates chip war with Japan

- US threatens \$300 million increase in tariffs
- Japan pleads innocence of agreement breach
- Third-market 'dumping' of chips continues

Washington

In what could be the first salvo of an open trade war, the United States last week announced that it would raise tariffs on as much as \$300 million in imports from Japan as a response to what are held to be violations of the agreement on semiconductor trade signed last summer. The tariffs will not be applied to semiconductors, but to other electronic products for which US companies still hold sizeable market shares.

President Ronald Reagan announced the sanctions on Friday 27 March, the day after his Economic Policy Council met to present him with options. Reagan accused Japan of failing to enforce "major provisions of the agreement" on semiconductors, and sided with US semiconductor manufacturers who have accused Japanese manufacturers of selling chips below cost to third countries for transshipment to the United States as a way around potential US duties. Such 'dumping' is prohibited by the trade agreement, but many agree it is difficult if not impossible to enforce. US manufacturers also say that they see no evidence of improved access to Japanese markets as promised by the semiconductor agreement.

Discontent over Japanese willingness to purchase US-made products was substantially augmented after a visit by a US trade delegation to Japan in February. Makoto Kuroda, vice-chief of MITI, the Ministry of International Trade and Industry, told the visiting delegation at a luncheon that attempts to sell US supercomputers to Japanese government agencies or universities would be futile. This comment caused a furore in US government circles, and apparently led to increased impetus on the government's behalf to retaliate against Japan.

A list of products to be covered by the tariffs will be published this week in the *Federal Register*. There will be a 14-day period for comment before a final list is announced. Tariffs will be as high as 100 per cent of import cost. A preliminary list includes televisions, radios, refrigerators, air conditioners, electric motors, pocket calculators and computer disk drives and central processing units.

At a meeting in Japan on 28 January, US officials warned that if third-market dumping did not stop in 30 days and market access improve in 60 days, the United States would "take appropriate steps to enforce the agreement". But Commerce Department data indicated that dumping

continued a month after the meeting.

US figures showed that dynamic random access memory chips were being sold at an average of 59.4 per cent below "fair price", and erasable programmable read-only memory chips were selling at 63.6 per cent below their fair price.

Congress seems ready to support the administration's retaliatory moves against Japan. On 19 March, the Senate passed by 93 votes to none a resolution calling on the president to take action on violations of the US-Japan semiconductor agreement. Members of the House of Representatives have indicated similar sentiments.

Recent Japanese efforts to reduce chip production and prevent dumping are described as "encouraging" by the president. But he is insisting that the trade sanctions will not be lifted until there is evidence that Japan is vigorously enforcing the semiconductor agreement. **Joseph Palca**



Japan's Prime Minister Yasuhiro Nakasone.

Japan dismayed by US retaliation

Tokyo

THE trade sanctions announced last week by the United States against Japan were greeted with dismay by Japanese officials, who insist that they have been adhering to the terms of the agreement on semiconductors signed on 31 July last year. The Ministry of International Trade and Industry (MITI) says that Japan has taken significant steps in the past few months both to encourage foreign chip imports and to tackle the root of what it sees as the problem — oversupply and rock-bottom prices.

The present problems in the chip trade can be traced back to 1984, when Japanese electronics companies invested massively in plant and equipment for memory-chip production. Oversupply and fierce com-

petition drove down prices, but Japan was able to capture 80 per cent of the world market.

The agreement was intended to protect the embattled US semiconductor industry as well as to provide access for US companies to the Japanese market. The provision in the agreement that chip prices in third-country markets should not be lower than those held to be fair for direct sales between Japan and the United States, was intended to prevent low-priced sales through intermediaries, but provoked many other companies to complain that they were being denied access to cheap Japanese supplies.

The immediate effect of the agreement in Japan was that prices plummeted whereas Japanese exports to third countries (in Latin America, South-East Asia and Europe) doubled as trading companies saddled with chips they could not sell to the United States unloaded their chips elsewhere at reduced prices. Many of these cheap chips have nevertheless found their way to the United States.

In mid-February, as US criticisms of Japan's failure to meet the agreement mounted, MITI instructed Japanese chip-makers to reduce production by 20 per cent in the remainder of the first quarter (January-March). But US complaints have continued unabated, and last week (23 March) MITI ordered a further 11 per cent cut in production in the next quarter.

Also, in early March, MITI set up the International Semiconductor Cooperation Center to promote the import of foreign chips, and Hajime Tamura, minister of international trade and industry, appealed to major Japanese semiconductor users to buy more foreign chips. But only Japanese users and one French company have so far joined the centre, even though the annual subscription fee is a mere Y200,000 (\$1,450). US exporters have shunned the centre on the grounds that it is controlled by their competitors.

According to MITI, Tamura wrote to US government officials on 28 March describing these developments and suggesting a joint US-Japan study of the third-market problem. MITI claims that the prices as low as \$1.8 a chip in South-East Asia are caused by Korean and even US producers selling at lower than economic prices so as to expand their market share. MITI also says that the supply of cheap chips from Japanese trading companies has now almost dried up. One MITI official has even suggested to the Japanese press that the recent disclosure of chip dumping in Hong Kong by the Oki Electric Industry Company may be the result of a subterfuge, or 'sting' organized by the United States.

David Swinbanks

NASA group urges space station changes to save space science

Stanford, California

FEARS that the continuing fallout from the Challenger accident will have a disastrous impact on space science over the next decade have prompted a NASA (National Aeronautics and Space Administration) science advisory group to recommend urgent new measures to get the space station in operation before 1995. In particular, the group recommends development of an unmanned heavy lift launch vehicle (HLLV) to launch space station structures, rather than relying solely on the hard-pressed space shuttle.

Events since the Challenger accident have shown that a "fundamentally revised plan of action" is required, according to a recent letter to NASA from Stanford University professor Peter Banks, head of the Task Force on the Scientific Uses of the Space Station.

The group's recommendations were prompted by "desperation", says Michael Wiskerchen, a Stanford senior research associate who is also a member of the task force. He says that steps must be taken now to save the next generation of space scientists and engineers. The committee expressed their concerns to NASA officials during a meeting at Stanford on March 19-20.

The committee recommends that, as well as development of the HLLV, NASA should investigate the possibility of launching a single large core-module to serve as both a laboratory and living quarters, and that it should extend the shuttle's ability to stay in space to two weeks or more.

The task force recommendations were initially received less than enthusiastically by NASA officials. Robert Naumann, chief scientist for the space station, responded to Banks saying that "if we wish to see an operational space station in our professional lifetime, the only option is to continue on our present course. Unless we all pull together on this, I'm afraid we will wind up with nothing."

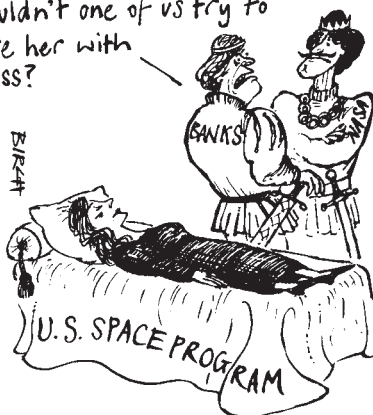
The task force supports the space station concept, but is worried about the effects of delays on the future of space science and is urging NASA to study its proposals in parallel with present plans for the space station so as to be able to take advantage of an HLLV if one is developed.

NASA recently completed a study of an HLLV, but has not received authority to develop such a vehicle; Naumann doubts that one could be operational by 1992. But members of the Banks committee say that private industry could have an HLLV ready by 1992. Banks's letter also says that access to the space environment is the "single most pressing issue of the next

decade" for the space science community. NASA's present plan to have the station operational by 1995-96 is "too late to help," he says.

While an extension of the duration of shuttle flights to 14-17 days from the present 7-day limit would "double the productivity of many missions . . .", Banks's advocacy of the development of an HLLV for launching elements of the space station stems directly from concerns about the

Shouldn't one of us try to wake her with a kiss?



length of time needed to assemble the station and to commission it if only the shuttle is available.

NASA's present plans call for a dozen or more shuttle flights to launch and assemble several smaller research and living modules. Wiskerchen said that the committee did not recommend that this plan should be abandoned, but that the large core module concept should be studied as well so that an informed choice between the two could be made if an HLLV programme is started.

Although there have been rumblings in the space science community that unmanned missions could accomplish most scientific objectives, Banks says there is still an important role for man in space, and that the proposed large-core module would permit laboratory-style research programmes that could not be done with automated systems. Lydia Dotto

• An Atlas-Centaur rocket launched from Cape Canaveral during a rainstorm on 26 March veered off course approximately 52 seconds after lift-off and was destroyed by ground controllers. The Atlas-Centaur was carrying a military communications satellite. NASA is seeking to recover the rocket's digital control unit to determine whether lightning may have contributed to the launch failure.

Initial investigations by NASA indicate that lightning was observed near the rocket at approximately the same time things started to go wrong with the flight. At 49 seconds after launch, the Atlas-Centaur's guidance system detected and

Indian rocket failure hits space programme

New Delhi

THE crash last week of a US\$100 million Indian rocket on its maiden flight and carrying a scientific satellite is a major setback to the country's space programme. The spacecraft went into the sea just two minutes after its launch from the Sriharikota launch pad and was witnessed by the Prime Minister, Mr Rajiv Gandhi.

The 40-tonne augmented satellite launch vehicle (ASLV) is believed to have fallen into the Bay of Bengal intact except for the two strap-on boosters, the only parts that seemed to work during the mission.

The four-stage ASLV was the first of India's second-generation rockets incorporating strap-on technology. It was also the first to use a closed loop guidance system for accurate ejection of payload in a predetermined orbit. The previous launcher, SLV-3, could lift only 50 kg of payload; with the addition of two boosters, carrying a total of 17 tonnes of solid fuel, ASLV's payload capability is 150 kg.

The prime minister was stoical after the failure, telling the team responsible not to be disheartened. "It is a small stumble, but work must go on", he said.

According to Professor U. R. Rao, chairman of the Space Commission, the mission failed after the strap-on boosters burned out and separated and before the ignition of the first stage. What happened subsequently was not immediately known as the telemetry failed. He said the exact cause of the crash would be known only after several days.

The rocket was intended to place a 150-kg satellite into a 400-km nearly circular orbit. The satellite carried scientific equipment to measure gamma-ray bursts. There is no plan to recover the satellite or the rocket from the sea.

The abortive launch is expected to delay other projects. India had planned to launch in 1989 a polar satellite launch vehicle using a liquid-fuelled engine, followed by a geostationary satellite launch vehicle using cryogenic engines in 1992. Last week's failure is likely to lead to a revision of the schedule. K.S.Jayaraman

then attempted to correct for a non-existent trajectory error. At 59 seconds, the launch vehicle began to break up, and ground controllers destroyed it. Although it was raining at the time of the launch, NASA officials say no lightning had been observed for the previous 12 minutes, and conditions were acceptable for launch.

This was the 67th Atlas-Centaur launch and the programme has had a 90 per cent success rate. The last Atlas-Centaur launch in December 1986 was a success, carrying another military communications satellite. Joseph Palca

Human genome sequencing plan wins unanimous approval in US

Gaithersburg, Maryland

THE project to map and sequence the human genome is now a big step nearer reality. At a meeting here last week, the Health and Environmental Research Advisory Committee (HERAC) of the Department of Energy (DoE) collectively took a deep breath and voted unanimously to approve in principle a draft report urging that \$1,000 million should be spent over the next seven years on the project. A final version of the report incorporating certain drafting changes is expected within four weeks.

Early goals for the DoE's efforts should be to make a physical map of the human genome. The first five years of the project would also see the development of new cloning and sequencing technology, sequencing of selected chromosomal regions and new means for handling data. The work would be done by DoE's national laboratories, as well as by universities and industry. But the HERAC subcommittee also says that there should be a place in the project for research proposals from single investigators as well as multi-disciplinary consortia. The committee recommends establishing two scientific panels, one to set goals and the other to review achievements.

Alvin Trivelpiece, director of the Office of Energy Research, asked HERAC to provide DoE with scientific advice on how and whether to proceed with "a potential major new initiative in human genome sequencing". Ignacio Tinoco of the University of California at Berkeley headed a subcommittee consisting of many of the leading lights in molecular genetics who wrote the report.

The energy department has already begun work on some of these objectives. The department has requested \$11.5 million in its 1988 budget for genome project activities. The HERAC report says that this figure should be doubled and that funds should increase linearly between now and 1993 to a plateau of \$200 million a year.

Charles DeLisi, head of the DoE's office of health and environmental research and a moving force behind DoE's efforts, says the decision to request additional funds rests with Trivelpiece's office. But Trivelpiece will be leaving DoE this month, and will be replaced at least temporarily by his deputy James Decker. Although characterized as an enthusiastic advocate of the project, Decker may not be willing or able to push a major new initiative for DoE. DeLisi may also leave DoE this year to take an academic position, but he believes the project has sufficient momentum to continue without him.

Congress is also a potential barrier to a

fully fledged mapping and sequencing project. Few in Congress will be enthusiastic about a new initiative that will cost more than \$1,000 million, although this is less upsetting than the estimate of \$3,000 million initially quoted for the project (see *Nature* 321, 371; 1986).

The HERAC report is at pains to explain why DoE should organize and administer the human genome initiative. Apart from the agency's long-term interest in the health effects of radiation, the draft report asserts that DoE has "unique capabilities to manage this initiative". While calling for cooperation among organizations involved, both domestic and international, the report nonetheless calls on DoE to implement its plans without delay and to take a lead in the project.

DoE's interest in becoming the lead agency does not necessarily sit well with other agencies that will be involved. James Wyngaarden, director of the National Institutes of Health (NIH), says it is presumptuous of an agency to claim a leadership role in a project on which it is at present spending less than \$10 million a year. Wyngaarden says NIH spend nearly \$300 million a year for research that deals at least in some part with sequencing and mapping human genes.

Wyngaarden is chairing a task force for the White House Domestic Policy Council looking at federal activities in the field.

Even as federal agencies debate the relative merits of the genome project, a private corporation has set out to beat everyone to the punch. Walter Gilbert of Harvard University has founded the Genome Corporation which, he says, will work out a cosmid map of the human genome for sale to anyone who may want it. He also intends to sequence segments of the genome, marketing a database containing that information.

Gilbert expects that customers will include the academic research community as well as the pharmaceutical industry. For the moment Genome Corporation exists only on paper, but Gilbert expects the company to be running by summer.

Gilbert will not speculate publicly on how long his mapping and sequencing effort will take, but admits that his timetable "is generally more aggressive than other people's". Gilbert also reckons the entire sequencing effort will cost far less than the DoE estimates — more like \$300 million — and could be accomplished within a decade by a modestly sized private company. Opinion is divided on whether Gilbert will be able to make a financial success out of the Genome Corporation, but his reputation forces most to take his efforts seriously. **Joseph Palca**

Soviets drill in Cyprus

London

Two research vessels, belonging to the Soviet Academy of Sciences, will shortly begin geological prospecting off southern Cyprus. The survey will form part of an international programme of research drilling into the Earth's crust and Cypriot scientists have been invited to participate.

Offshore resources are a delicate matter in the eastern Mediterranean and Aegean (see *Nature* 326, 233; 1987), because of conflicting claims to territorial waters and the possibility of fossil fuels in the area. In the case of the forthcoming Soviet expedition, the official announcements are somewhat conflicting. On 12 March, the Soviet Ambassador in Nicosia, Yuri Fokin, said that the main purpose of the operation would be to map the "complicated and unique" geological structure of the deep-water canyon of the eastern Mediterranean at a depth of 10–15 km, and said that at this depth there is evidence that could indicate the presence of oil or natural gas. But the following day an unnamed government spokesman stressed that these explorations are not aimed at oil prospecting. **V.R.**

Beer gut prize is awarded

London

TOMAS Hokfelt and Viktor Mutt, both professors at the Karolinska Institute of Stockholm, share the 1987 biennial Artois-Baillet Latour Health Prize for their work on neuropeptides. In fact, 64-year-old Mutt has worked primarily on gut peptides, many of which have later been shown to be neuropeptides, whereas 47-year-old Hokfelt has concentrated on central nervous system peptides. Given every other year for research with immediate applications, the £83,000 prize is awarded by a Belgian foundation to which Count Alfred de Baillet Latour donated his shares in Artois Breweries six years before his death in 1980. **P.N.**

Polish journals cutback

London

POLAND'S economic problems continue to make it impossible for Polish scientists to keep up with foreign developments "on a day to day basis", the Polish Technical Writing Council has noted. At a meeting in Warsaw in March, the council said that "drastic limitations" on hard currency for journal subscriptions and foreign books will prevail. Moreover, shortages of paper and inadequate and outdated printing technology mean that even Polish learned journals have had their print-runs and size reduced, and frequently suffer considerable delays in production. Much subject matter is therefore outdated before it appears, and the "substantive and editorial standards" of these journals have declined considerably. **V.R.**

Moscow and Bonn unite on environment

Munich

WEST Germany hopes "soon" to sign a broad agreement with the Soviet Union on the protection of the environment, according to the West German Environment Ministry last week. The announcement was made at the culmination of five days of talks with Soviet officials in Bonn.

The two-year pact will cover the prevention of air, water and land pollution as well as the prevention of accidents harmful to the environment. In addition, scientific and industrial exchanges between the two countries will also be outlined. The details of the agreement will be released only when it has finally been signed.

The environmental pact is just one indication of a thaw in relations between the two countries. West German cabinet ministers, including Research and Technology Minister Heinz Riesenhuber, are scurrying to set up appointments in Moscow in April, even before West German President Richard von Weizsäcker pays a newly scheduled visit there in May.

Riesenhuber is expected to sign an agreement on the peaceful uses of nuclear energy which has been on ice since late last year. The Soviets cancelled a visit by Riesenhuber to Moscow, planned for December, because of remarks made by German Chancellor Helmut Kohl in November, comparing Soviet leader Mikhail Gorbachev with Hitler's propaganda minister during the war, Josef Goebbels.

Two other agreements, on medical and agricultural research, are also expected to be signed by the two sides within the next few months.

The new pact with the Soviet Union may help to give the parallel negotiations with East Germany a stimulus, according to a source in the Environment Ministry. A high West German official, Wolfgang Schäuble, returned on Friday from talks with East German leader Erich Honecker in East Berlin, saying that significant progress had been made on East-West German relations. The two-day meeting may have included a decision to sign (finally) a scientific agreement between the two German states which has been under negotiation since 1973. But Schäuble refused to discuss the details of the talks, which also dealt with the invitation of Honecker to visit West Berlin on 30 April for the opening of the 750th anniversary celebration.

Steven Dickman

AZT given the green light for clinical treatment of AIDS

Washington

THE US Food and Drug Administration (FDA) announced on 20 March the approval of the first drug developed to combat AIDS (acquired immune deficiency syndrome). The drug zidovudine, commonly known as azidothymidine (AZT), will be marketed by Burroughs Wellcome under the trade name Retrovir.

AZT passed through human trials and cleared the FDA in 22 months, under an accelerated review process set up to speed the approval of breakthrough drugs. To receive AZT treatment now, AIDS patients must be stricken with *Pneumocystis carinii* pneumonia or exhibit a CD4 (T4) lymphocyte count of less than 200. Patients with less severe AIDS and ARC (AIDS-related complex) indications may receive AZT once adequate supplies are assured.

Until production can meet demand for the drug to treat the approximately 14,000 AIDS cases and 40,000 advanced ARC cases now known in the United States, Burroughs Wellcome has set up a Retrovir

control centre in Washington, DC to serve as a clearing house for applications submitted by physicians for individual AIDS patients. The centre will authorize shipment of the drug based on the type and severity of the patient's symptoms.

AZT disrupts the lifecycle of the AIDS virus by blocking the action of the enzyme reverse transcriptase (see *Nature* 325, 773; 1987). Although not a cure for AIDS, AZT improves the short-term survival of AIDS patients. But it does not come without side-effects, such as severe anaemia, and its long-term effects are unknown.

According to Burroughs Wellcome, treating one AIDS patient with AZT will cost about \$8,000–10,000 (US) a year. Representative Henry A. Waxman (Democrat, California) has asked the House Budget Committee for \$60 million in 1988 to help needy AIDS patients shoulder this financial burden. He has also appealed to the Appropriations Committee for emergency funding for those who cannot afford the drug this year.

Carroll Ezzell

Inquiry cleared Gagarin of pilot error

London

YURI Gagarin, the first man in space, who died in a training flight in March 1968, has finally been cleared of suspicion of pilot error and/or being at the controls of an aircraft while drunk. Although no such posthumous accusations were ever brought openly against Gagarin, and he remained officially a member of the Soviet space pantheon, the findings of the official inquiry into his fatal crash were never made public. This inevitably produced rumours about the causes of the accident. According to Valentina Tereshkova, the first woman cosmonaut, in the mid-1960s, the Soviet Union was planning a space mission to be led by Gagarin. No such mission materialized, fuelling speculation that Gagarin was out of favour.

Last week, however, *Pravda* belatedly published an account of the official inquiry into Gagarin's last flight, signed by Professor S. Belotserkovskii, winner of a state prize for his work in space science, and cosmonaut Aleksei Leonov. They concentrate on two main issues: the "aviation technology" of the aircraft and a "complex of questions" including pilot readiness, flight plan and adherence to safety regulations.

The commission established that all on-board systems operated faultlessly up to the very end. There was no evidence of fatigue in any of the components recovered nor of on-board fire or explosion. The

engine was operating until impact, the electrical system and oxygen supply were working correctly, and the pilot's radio was on. Neither Gagarin nor his co-pilot Vladimir Seregin tried to eject.

Evidence from fellow-cosmonauts, who attended the pre-flight briefings with him on 26 and 27 March, and from doctors who analysed the tape-recordings of his final flight, indicated that Gagarin's behaviour and voice patterns revealed no unusual stress.

What, then, went wrong? According to the flight plan, the aircraft should have been making a turn from 70° to 320° while gradually descending between two layers of almost continuous cloud. Instead, it suddenly went into a near-vertical dive. Possible explanations are suggested: an "unexpected obstacle" (another aircraft, an aerosonde balloon, or possibly a flock of birds); encountering the slipstream of another aircraft; or hitting a rising vertical airstream could increase the angle of attack sufficiently to produce a stall. Whatever the cause, Gagarin and Seregin made a determined effort to fight their way out of the crisis. On the basis of the flight data and additional evidence of the trajectory, including the ground and tree-tops at the impact site, it appears that they almost succeeded. It has taken nineteen years and three changes of leadership for the Soviet people to be permitted to know this.

Vera Rich

Research council suggested as boost to Australian science

Sydney

AN Australian Research Council should be established as a channel for the federal government's direct support of basic research. This is the chief recommendation contained in a report from the Australian Science and Technology Council (ASTEC) on research in higher education, now before parliament. The council took 15 months to produce the report and considered 426 submissions.

The recommendations are intended to improve Australia's return on its investment by encouraging more research that will contribute directly to new products and processes as well as to improve the yield on that investment by persuading universities to specialize in areas where they have special expertise.

The report claims that supporting Australian tertiary institutions primarily for teaching rather than research has prevented the efficient use of research funds. Most funds for university research are now channelled through the Commonwealth Tertiary Education Commission (CTEC), but as part of the recurrent funds for university support, are not earmarked as such. The assumption is that university academics spend 30 per cent of their time on research and the rest teaching. Because staff numbers are based on the numbers of students, funds are spread evenly across all disciplines and, ASTEC says, Australian universities therefore end up as carbon copies of each other.

To ensure a greater concentration of research resources, ASTEC has called on CTEC to play a greater role in the distribution of funds within individual institutions, which have been asked to develop and publish their research strategies. In 1984, CTEC provided A\$643 million for research in universities, which is comparable with the level of support in most other countries of the Organization for Economic Cooperation and Development, but A\$339 million of it corresponds to the salary equivalent of the time that academic staff spend on research, leaving just A\$86 million for salaries of technical and administrative support staff, equipment and expendable research materials.

On this basis, ASTEC concludes that specialization is the best way to produce significantly better results, even though this will mean that some academics will find themselves doing less research and more teaching. But university staff associations hold that every university academic has the right to do research and will bitterly oppose moves to create a subdivision of university staff who are allowed only to teach.

The Australian Research Council

(ARC) would be a statutory body answerable to the minister for science and would replace present schemes for direct research support, including the Australian Research Grants Scheme (ARGS) and the Commonwealth Program for the Promotion of Excellence in Research (see *Nature* 326, 5; 1987). Their combined budgets would provide the ARC with A\$46 million. ASTEC estimates that the ARC will need around A\$ 96 million a year. It recommends that it be given A\$66 million in 1987-88 and built up to the full amount over the next four years. This would bring the ARC budget, as a fraction of gross national product and total higher education research spending, to a level comparable with research councils in other developed countries.

The study also recommends that the extra money should not be taken from CTEC's allocation. But in a hard economic climate where a 10 per cent increase (A\$3 million) in funds for the ARGS in the 1986 budget was hailed as a real victory for science, implementation of the last recommendation is unlikely.

Charles Morgan

US-Soviet test-site monitoring a success

London

THE first stage of the joint Soviet-US experiment to monitor underground nuclear tests has ended successfully, a member of the US team, Professor Jonathan Berger of the University of California told *Izvestiya* last week.

The experiment, which is to run until July 1987, is sponsored jointly by the Soviet Academy of Sciences and the (non-government) US Natural Resources Research Council. The purpose of the experiment is to demonstrate that international verification of an underground test-ban treaty would be technically feasible, and that state-of-the-art seismic monitoring equipment is sufficiently accurate to eliminate erroneous readings from background noise or cavity effects.

The second stage of the experiment, Berger says, involves the siting of US monitoring equipment in 100-m deep boreholes at a site near the Semipalatinsk test-site in Kazakhstan, which, Berger says, have already been drilled. A similar network of boreholes is also planned in Nevada and California, to monitor the US test-site. But although these monitoring stations are now almost ready, the Soviet team have not yet received clearance from Washington to visit them. Vera Rich

EEC budget row threatening the research worker

London

THE jobs of 3,000 European research workers engaged on collaborative research projects could be lost if the members of the European Economic Community (EEC) cannot agree on a research and development budget for the next five years. Last week the ministers failed, yet again, to agree the budget for the research programme, Framework, which was supposed to be effective on 1 January this year.

West Germany and Britain have been given until the end of this week to agree, although it remains unclear what sanctions will be imposed if they do not comply. Initially these two states, in partnership with France, challenged the proposed £5,390 million budget.

France relented but West Germany and the United Kingdom remain in opposition. Britain favours a ceiling of £2,940 million, says the Community must seek better value for money, use up 1,000 million ECU from the last budget and have a more professional approach to managing research projects.

The ministers have been trying to agree since last autumn when the EEC proposed its budget. By last week an unofficial compromise of £4,600 million was being discussed but still without success.

On his return from the European debate, Mr Geoffrey Pattie, Minister for Information Technology, scathingly criticized the EEC's management of information technology research, referring to it as a shambles. But the EEC remains determined that the new budget should be approved by all member states and has been given the backing of the European parliament.

The predictions of possible job losses came from members of the European Parliament (MEPs) who reject claims of member states that they can conduct effective national research and development programmes to compete against the might of Japan and the United States. And member states backing the new research programme maintain that the governments which oppose the Commission's proposals have never indicated which programmes they wish to cut back or cancel.

Three thousand researchers on information technology projects will have their contracts suspended "Because the adoption of the specific programmes is dependent on that of the general programme", claim the MEPs. "The whole confidence of the European scientific community would be shaken."

Bill Johnstone

Towards room-temperature superconductivity

The past few weeks have witnessed an astonishing preoccupation with the hitherto esoteric and inaccessible phenomenon of superconductivity. To banish electrical resistance may be the contemporary equivalent of the philosophers' stone. Nature correspondents say how it seems to them.

German ambivalence

Munich

WEST Germany is in a curious state of confusion about oxide superconductivity. Most people are impressed that the first dramas were acted out at the IBM laboratory at Zurich, where the tunnelling electron microscope won a Nobel Prize for its inventors only a year ago. Many of those working in the field are away this week, still on the way back from the American Physical Society's meeting in New York. The meeting of the German Physical Society earlier in March appears not to have satisfied all curiosity.

For the time being, however, the Technische Hochschule at Darmstadt seems to be in the lead. Frank Steglich and his group have been looking at the feasibility of making Josephson junctions of the new materials; their tentative conclusion is that the boundaries between grains in the ceramic materials are responsible for most departures from superconductivity, which is an incentive to make materials with bigger crystals.

The Nuclear Research Centre at Karlsruhe supports two active groups, as does Munich. Both centres are playing with the substitution of one rare-earth element for another. Karlsruhe is eager to measure electrical properties. The University of Giessen is also active.

Most workers in the field are surprised at the speed with which events are moving. Professor Saemann-Ischenko at the University of Erlangen is struck by the advance towards industrial applications of the new materials. Steglich, at the University of Darmstadt, says that West Germany was "a little bit behind" at the beginning, but that the field is still wide open.

Steven Dickman

Japanese gird loins

Tokyo

AN heroic effort is under way here to develop high-temperature superconductors before the United States.

At the semi-annual meeting of the Physical Society of Japan in Nagoya on 28 March, an afternoon session on high-temperature conductivity was swamped with presentations by more than 50 speakers from government, industry and the universities. Talks were strictly limited to 10 minutes, including discussion, and peo-

ple jammed the aisles and spilled out of the doors of the conference hall as they jostled and craned to hear the proceedings, which continued well into the night under the glare of television camera lights.

It is less than four months since a group of researchers at Tokyo University confirmed superconductivity at about 30 K in a copper oxide ceramic. But research is already under way on a broad front. Some speakers talked on crystal structure and theory, others reported more practical findings: researchers from NEC Corporation presented evidence of the alternating current Josephson effect in a yttrium-barium oxide at about 90 K; Nippon Telephone and Telegraph (NTT) reported superconductivity at 90 K in a new lanthanum-barium-copper oxide, an important discovery as lanthanum is much cheaper than yttrium; and Fujikura Densen, an electric cable manufacturer, showed pictures of a crude loop of wire fashioned from the new ceramics.

"I used to play golf every weekend until this happened", lamented Professor Shoji Tanaka, leader of the Tokyo University group, as he took a breather for coffee during the marathon session. Now Tanaka's time is fully occupied shuttling back and forth between meetings and conferences both inside and outside Japan.

Tanaka is particularly excited by his group's discovery that ceramics doped with paramagnetic rare-earth elements (such as erbium and holmium) rather than yttrium retain their superconductivity and have similar critical temperatures to the yttrium ceramics. He thinks this finding may lead the way to ceramics with high critical currents, thereby opening up a treasure chest of applications, including superconducting magnets, energy storage devices, superconducting computers and cheap magnetically levitated trains (the Nagoya session happened to be held at the same time as the Japan National Railways carried out the first test run of a prototype train levitated by conventional superconducting magnets — the train cost about \$10 million (¥1,500 million) to build).

But Tanaka worries that the race to develop such applications may lead to a conflict with the United States, similar to problems now seen in the chip and supercomputer trade. IBM and most other US companies drastically scaled down their research on Josephson junctions in 1983,

partly because of the complications of cooling with liquid helium, but Japanese companies such as Hitachi and NEC have forged ahead. Last year, Japan's Research and Development Corporation awarded a \$10 million grant for five years to Professor Goto of Tokyo University to develop a superconducting magnetic flux computer under the ERATO programme (see *Nature* 324, 507; 1986).

A member of the editorial board of the *Japanese Journal of Applied Physics* echoed a similar sentiment to Tanaka when he announced during the Nagoya session that he, not an American, would correct the English of the flood of papers the journal had received. Although said as a joke, it does reflect the keen sense of competition with the United States.

Government ministries have also been quick to spot a potential winner. The Science and Technology Agency (STA) set up a committee in late February, drawn from universities, industry and government laboratories, to organize symposia, analyse research trends and publish reports on the new superconductors. Funding for the committee's activities comes not from the government but from a group of about 40 companies, including electronics giants, mining concerns and electric cable manufacturers, who have contributed ¥200,000 each to the fund.

A similar committee has been set up by MITI, the Ministry of International Trade and Industry, although, according to Tanaka, there are more members from industry than in the STA committee (Tanaka serves on both). MITI is also considering applying the new superconductors in its energy-saving Moonlight project, for example, by developing energy-storage devices.

Last year, the Ministry of Education, Science and Culture thought that a special project on superconductivity headed by Professor Nakajima of Tokai University (formerly of Tokyo University) was a failure. Now the ministry has decided to extend the three-year programme by an extra year. But, apart from this small commitment, no major government funds have yet been assigned to research on the new superconductors. Undoubtedly the bulk of the drive and support for research will come from industry. And judging from Nagoya, research in that sector is already well ahead.

David Swinbanks

Mild enthusiasm

Washington

THE enthusiastic attendance at the late-night meeting of the American Physical Society, in New York on 19 March, was proof enough of people's determination to break the record in the race for the superconductor with the highest transition temperature. But in the absence of a simple explanation of the new results, researchers are relying on inspired guesswork

and empirical extrapolation in the design of new materials.

Most people appear however to agree that, for the time being, the starting point for the new wave of excitement is the discovery (at the IBM laboratory at Zurich, where tunnelling electron microscopy was first made a reality) that lanthanum copper oxide can be superconducting. Since then, yttrium has been successfully substituted for lanthanum, as have been lutetium (at Brookhaven) and gadolinium (at the Sandia Laboratories).

Dave Ginley of Sandia makes no bones about it: it is possible to make a whole series of crystalline structures by substituting one rare earth for another. The small variations of crystal structure caused by the different ionic radii of the constituents yield solids with different properties. The trick, for the time being, is to ring the changes that work in the right direction, that of a higher transition temperature.

Meanwhile, more and more attention is being paid to techniques of fabrication. The current-carrying capacity of the materials so far made is no more than one ampere per mm^2 , rather less than is usually carried by a domestic power cable. Jim Smith, director of the Center for Materials Science at the Los Alamos National Laboratory, noting that the materials so far constructed are little better than "compressed powders", argues that the future must lie in more careful heat treatment of the ceramics to yield materials more like glasses that will carry a higher current density.

Meanwhile, much attention is being paid to techniques for making flexible stranded cables out of brittle ceramics. Much of the technology of the classical superconductors turns out to be relevant: niobium-titanium and niobium-tin are hard and glassy in themselves, but can be extruded into fine flexible wires when embedded in copper. Curious as it may seem, nobody working in this new field is dismayed at the prospect of putting the new materials to large-scale use. The question is when, and at what temperature.

David Lindley

Late starters

London

BRITISH research groups are working fast to catch up with their Japanese and US counterparts in the race to exploit the new superconducting ceramics. But many are worried that cuts in grants have undermined their efforts before they start. In just two months, researchers have switched from long-standing projects to the novel topic; all the groups that have declared themselves have succeeded in synthesizing and characterizing the compounds reported by Chu in Houston and other workers.

At a meeting last week of the Rare Earth and Actinides group sponsored by

the Science and Engineering Research Council (SERC) at Birmingham, three progress reports showed the strides taken in the past weeks. A collaboration between the Universities of Warwick, Durham and Cambridge and the Rutherford Appleton Laboratory (RAL) has succeeded in characterizing the crystal structure of the lanthanum-barium ceramic (LaBaCuO_4) using the ISIS neutron facility at RAL. This high-resolution facility gives them, they believe, a unique advantage over groups elsewhere, and has revealed fine alterations in structure and phonon characteristics that coincide with changes in electrical behaviour.

The University of Birmingham boasts another collaboration, all on one site, using various skills in sample preparation and low-temperature work already available. The researchers there have measured the magnetic flux through a cylinder of the yttrium-barium ceramic (YBaCuO_4) and shown that it is quantized, which confirms the superconducting nature of the material. The value of the quantum is not yet certain, but their result means that there must be persistent currents, and they say that the resistance of the sample is less than 10^{-11} ohms.

Other groups not represented at the meeting have also made rapid progress. At Cambridge, several groups are involved. They are relying on proven expertise in the manufacture of high-purity single crystals to obtain samples that will allow them to investigate the microscopic properties that cause the superconductivity. With a history in the preparation of the related ceramics LaNiCuO_4 , the department of metallurgy at Imperial College, London, is working with the physics department, where they have long experience in composite superconductors. They are particularly interested in the effects of grain boundaries in these compacted materials. Research is also under way at Oxford and Bristol.

All those involved are worried that the squeeze on university budgets has left them in a weakened position. One says that the new opportunity is also an example of the effect of the drive for efficiency in universities, in which the nature of long-term research has been forgotten. Compared with companies such as AT&T Bell and IBM, which can put large numbers into the field at short notice, Dr Chris Muirhead of the University of Birmingham believes only a handful of researchers can become involved in British universities. The present consensus is that staffing and not equipment is the problem — most laboratories are using existing equipment, sometimes hastily adapted. The general view is also that British research is best concentrated on fundamental physics and not on the race for higher temperatures.

SERC can offer little more than sym-

pathy. The last round of grant applications came before the new potential was realized. Now £8 million overcommitted, the next round of SERC grants is virtually cancelled except for specific initiatives into which superconductor research may be slotted. According to John Farrow, secretary of the SERC physics committee, the committee recognizes "that this research must be supported urgently"; he says that a case for supporting the research out of the '21st Century Materials Initiative' is being put together. But in the present round, he expects support to be in the region of £200,000 — enough to support one or two key groups. He says that the council expects to be able to support larger applications by about September.

SERC has not yet promoted any combined university/industry research programme, but it is considering setting up a joint meeting of interested parties. The LINK scheme of the Department of Trade and Industry (DTI), designed to stimulate such cooperation, includes a programme for electronic materials, but the relevant committee has yet to meet. A spokesman for the DTI said that there was also the opportunity for collaborative schemes under the Support for Innovation project, citing materials development as an appropriate area. He went on to say that "this work is still at the basic research level, not with any specific applications at the moment", so that DTI involvement was not necessarily appropriate just now. Given the enthusiasm for the new opportunity at the New York and Nagoya meetings, it may be that British companies will be left behind.

Roland Pease

Anything goes?

London

IN the race to reach the record temperature for the onset of superconductivity, almost anything goes. A field developing as rapidly as this is fertile ground for rumour, and it is often difficult to distinguish fact, tentative result and speculation. The *China People's Daily* of 19 March reported a spectacular advance with the onset of superconductivity at 215K, though the transition is said to be very broad, with a midpoint at 93K.

If true, this work from the Chinese Academy of Sciences represents a substantial 100K advance on previously reported records. Still more remarkable is the claim from Dr Kozo Ohara's group at Kagoshima University in Japan of the possibility of superconduction at 14°C — room temperature. This result, originally published in the Japanese daily newspapers on 18 March, was hotly debated at the recent meeting in Nagoya (see above) as zero conductivity was not measured, although other indicators of superconductivity (the Meissner and Josephson effects) were. Experts at the conference said it was "too soon to call this superconductivity". R.P.

Undermining Australian parks

SIR—Why are the Northern Territory government, mining industry and certain conservationists so strongly opposing the application of the federal government of Australia for the World Heritage listing of Stage 1 of Kakadu National Park to be extended to Stage 2? The reasons are complex. In his article on "Politics and wildlife in Australia" (see *Nature* 325, 112; 1987), Kenneth Mellanby claimed that Stage 2 of Kakadu "must be the most boring national park in the world" and that it is "substandard" for consideration as a World Heritage site. He and the Australian media conservationist Harry Butler have likened it to a "clapped-out Holden" while members of the Northern Territory government and NT News frequently like to refer to it as "clapped-out buffalo country".

In support of this claim, the Northern Territory government commissioned a video film, in which Harry Butler and Kenneth Mellanby deprecated Kakadu Stage 2. This film depicted a biological wilderness where large herds of feral buffalo were claimed to be responsible for severe environmental degradation. Signs of over-grazing, trampling, wallows and buffalo swim channels were all shown. Certainly buffalo can cause considerable environmental damage. Swim channels in the wetlands may have been one cause of saltwater intrusions into the floodplains, and overgrazing by buffalo has resulted in substantial changes in the vegetation and some erosion in the past. The vegetation has, however, shown a remarkable recovery to its former status since the Australian National Parks and Wildlife Service instigated an intense control programme of the feral buffalo.

The recovery of the vegetation has been rapid in both the woodlands and wetlands. Scientists of CSIRO's Tropical Ecology Research Centre at Darwin have been monitoring these changes. Comparison of a fenced-off area containing buffalo with others vividly illustrates the dramatic changes that have occurred both in the physical structure and species composition of the vegetation since the removal of the buffalo. Yet it was in this maintained high density area of buffalo that Mellanby and Butler did their filming. So misleading was the video that the Darwin group countered this picture of Stage 2 based on their own studies. Why do the two sides differ? Inevitably the judgement of landscapes and biotas is to a certain extent subjective, but, as Paul Adam remarks (*Nature* 325, 570; 1987), to describe Stage 2 of Kakadu as "the most boring national park in the world" is a greater reflection on the writer than on the area. Nevertheless the views of the mining industry and the Northern Territory government coincide with those of their well-paid ecological consultants.

What is really at stake here is not the World Heritage value of Kakadu Stage 2 but the control of the park and the mining of uranium. Anything that is going to make mining more difficult is clearly against the interests of the mining companies and can be expected to be opposed by them. There is already one uranium mine (Ranger) that is technically not within the confines of the park but which is entirely surrounded by it. Most of the other known mineral resources (and these are considerable) are in similar positions: excised from the park but entirely surrounded by it.

In the Australian National Parks and Wildlife Service Report to the Senate Standing Committee on National Resources, it was estimated that only 3 per cent of the total mineral resources in the area fell within Stages 1 and 2 of Kakadu National Park or the proposed Stage 3. Nevertheless, decisions to mine within these areas cannot fail to be affected by the conservation value of the surrounding land. Although technically outside the park, they cannot be seen as separate from it. Ranger uranium mine is, for example, now applying to release water into the park when previously it had agreed not to do so.

The exploitation of the Jabiluka mine lease worth an estimated A\$19,000 million could also pose serious environmental problems as it occurs within an area that is seasonally inundated. For the present, the Australian Prime Minister, Bob Hawke, has stated that he wants to see no further mining within the Kakadu area, but a change in government or economic circumstances could well lead to a change in this decision. The exploitation of these mineral resources will then clearly be a lot easier if they are not surrounded by a World Heritage Conservation site.

Finally, what can be said of the views of the aborigines, the traditional owners of much of the land within Stages 1 and 2 of Kakadu? Mellanby gives the impression that the park authorities are out of touch with their views. My experience was entirely to the contrary. There is considerable understanding and liaison: moreover, 40 per cent of the Australian National Parks and Wildlife Service staff are aborigines. They receive royalty payments from the uranium mining at Ranger but they, like other Australians, are divided on whether further exploitation should take place. The considerable pressures that have been put on them to accept mining have, however, been eloquently expressed by their leader, Bill Neidjie, in a recent Australian television report.

The mix of ecological, economic, social and political factors represented in the Kakadu debate are complex. The conser-

vation and mineral wealth of the area are both extremely large. Legal action in the federal courts has led to a postponement over any decision on the World Heritage listing of Kakadu Stage 2. But whatever the decision, the battle over mining is likely to continue: the financial rewards are too great for it to be otherwise.

ANDREW WATKINSON

*School of Biological Sciences,
University of East Anglia,
Norwich NR4 7TJ, UK*

Food for thought

SIR—Some archaeologists suffer from an unnecessary anxiety that may affect their perception of the evolution of early man. On the one hand, it has recently become apparent that meat-eating is a health hazard for our species. On the other, our immediate forebears, the *Homo sapiens* evolving in periglacial environments, must have been substantially carnivorous. Surely evolution should have fitted modern man to a diet at least as rich in animal material as his present one?

A popular explanation is that lack of exercise accounts for the condition of modern man. But exercise alone does not seem to be sufficient to protect us from the cardiovascular ailments consequent on meat-eating.

An explanation that appears to have escaped attention is that modern prey species may be to blame. Wild deer, ibex and so on yield exceedingly lean meat, low in fat, quite unlike the meat consumed in modern diets. Modern animals are deliberately denied the benefits of aerobic exercise.

Our health problem, it would seem, can be overcome by the straightforward expedient of avoiding being placed higher in the food chain than a lazy animal.

GEOFFREY KING

*1267 Four Mile Canyon Drive,
Boulder, Colorado 80302, USA*

ALLAN LINDH

*4185 Page Mill Road,
Los Altos Hills,
California 94022, USA*

Aluminium alarm

SIR—I read with interest of the concern of Coriat and Gillard (*Nature* 321, 570; 1986) about aluminium in tea, and the report of Tennakone and Wickeramanayake (*Nature* 325, 201; 1987) on aluminium from cooking utensils.

Why are they concerned about aluminium from these sources compared to the substantially larger amounts added directly to many foods, such as alum in baking powder and pickles?

G. M. WILD

*Eli Lilly and Company,
Lilly Corporate Center,
Indianapolis, Indiana 46285, USA*

The chronology of AIDS research

This statement from Gallo and Montagnier constitutes part of the agreement between the US and French AIDS research groups.

WE set out a brief chronological history of some critical published facts, in the period up to May 1985, on the discovery and demonstration of AIDS as a retroviral disease.

Retroviruses

1970–73. Following Temin's hypothesis that RNA tumour viruses replicate via a provirus DNA intermediate, Temin and Mizutani (1970)¹, and independently Baltimore (1970)², discover reverse transcriptase. Later, the existence and integration of infectious proviral DNA is demonstrated by Hill and Hillova (1971)³ and confirmed by Montagnier and Vigier (1972)⁴, Svoboda and co-workers (1973)⁵, and others. Spiegelman (1970)⁶, Gallo (1971, 1972)^{7,8}, Gerwin *et al.* (1972)⁹ and others independently develop useful sensitive specific assays for reverse transcriptase of retroviruses.

1976. Morgan, Ruscetti and Gallo (1976)¹⁰ discover T-cell growth factor, or interleukin-2 (IL-2), necessary for long-term *in vitro* cultivation of human T cells.

1979. Barre-Sinoussi, Montagnier, Lideau, Sisman, Wood and Chermann (1979)¹¹ show that antibody against alpha interferon allows significant increase in mouse retrovirus production by infected cells.

Human retroviruses

1980–82. Gallo, Poiesz and co-workers (1980) isolate¹² and (1981) characterize¹³ the first human retrovirus, called human T-cell leukaemia virus type I (HTLV-I). Hinuma *et al.* (1981)¹⁴ identify C-type virus in a cell line isolated by Miyoshi from a patient with adult T-cell leukaemia (ATL) and detect antibodies against antigen-bearing cells in patients with ATL. Yoshida and co-workers (1982)¹⁵ isolate and characterize adult T-cell leukaemia virus (ATLV) and demonstrate clonal integration of ATL proviral DNA in leukaemic cells of patients. In a cooperative study, Gallo and co-workers together with Miyoshi and co-workers (1982)¹⁶ show ATL to be identical with HTLV-I. In collaborative studies, Catovsky, Blattner, Gallo and co-workers (1982)^{17,18} show HTLV-I to be a contributing aetiological factor in T-cell leukaemia in the Caribbean region. Gallo and co-workers (1982)¹⁹ isolate a second type of retrovirus, HTLV-II, from a cell line obtained from a patient with hairy-cell leukaemia.

AIDS

1981. Gottlieb and co-workers (1982)²⁰, Friedman-Kien and co-workers (1981)²¹, Siegel and co-workers (1981)²², Masur and

co-workers (1981)²³ and Mildvan and co-workers (1982)²⁴ independently diagnose a new disease, AIDS, in groups of young homosexual men.

1982. Epidemiological evidence suggesting that AIDS is a new infectious disease is developed by the Centers for Disease Control (1982)²⁵.

February 1983. At the Cold Spring Harbor Workshop on AIDS, Gallo proposes that AIDS is probably caused by a retrovirus, presumably a variant of HTLV-I or II.

May 1983. Barre-Sinoussi, Chermann, Montagnier and co-workers (1983)²⁶ publish: (1) the isolation and identification of a non-transforming retrovirus (later called lymphadenopathy-associated virus

HTLV-I variant in 2 of 33 AIDS patients. **September 1983.** At the Cold Spring Harbor meeting on human T-cell leukaemia-lymphoma viruses:

Montagnier and co-workers (1984)²⁹ report: (1) the identification of LAV-like viruses from 5 patients with lymphadenopathy and 3 patients with AIDS (homosexual, haemophiliac, Haitian); (2) the selective affinity of LAV for CD4 (T4) helper lymphocytes; (3) the presence of antibodies (enzyme-linked immunosorbent assay, ELISA) against the main LAV antigens in patients with lymphadenopathy-associated syndrome (LAS) (63%) and AIDS (20%); (4) that LAV is morphologically similar to equine infectious anaemia virus (EIAV) and different

This history is by no means exhaustive. It outlines the main contributions of the two central parties and of some other groups to the determination of the causative agent of AIDS (acquired immune deficiency syndrome). This is not to diminish the important contribution made in this field (either independently or in collaboration with us) by other laboratories and clinicians distributed worldwide.

Both sides wish it to be known that from the beginning there has been a spirit of scientific cooperation and a free exchange of ideas, biological materials and personnel between Dr Gallo's and Dr Montagnier's laboratories. This spirit has never ceased despite the legal problems and will be the basis of a renewed mutual cooperation in the future.

The parties also believe that some clarification is needed regarding their position on nomenclature. Various generic names have been given to the AIDS virus. Dr Gallo and his collaborators named it HTLV-III in their May 1984 publications according to a recommendation made in September 1983 at the Cold Spring Harbor meeting on HTLVs by a group of ten European, Japanese and American retrovirologists. They suggested that names of human retroviruses discovered in the future related to but distinct from human T-leukaemia virus be named sequentially, numbered HTLV-III, IV, and so on.

Dr Montagnier and his collaborators first reported in May 1983 the identification of a novel human retrovirus not closely related to HTLV-I and II and in public meetings from September 1983 and subsequently in publications from 1984, they named this virus LAV for lymphadenopathy-associated virus because the first patient from which it was isolated had lymphadenopathy syndrome.

The US Department of Health and Human Services officially assumed a double generic name HTLV-III/LAV in recognition of the contributions of both sides.

When the genome structure of the AIDS virus was determined in 1985, the results showed some major differences in the organization of the genetic information from HTLV-I and II. This prompted another and more formal nomenclature committee to suggest a simpler and new generic name as the virus belonged to an entirely new class of human retroviruses. Their suggestion was to use the generic term HIV — human immunodeficiency virus. They also suggested maintaining specific strain names (LAV-I_{BRU}, LAV-I_{LOI}, HTLV-III_B, HTLV-III_{RF}, ARV1...) in the interest of continuity and appreciation of the biological differences of the strains. Both groups agree with this recommendation.

(LAV)), different from HTLV-I and HTLV-II, in cultures of T lymphocytes derived from a patient with lymphadenopathy syndrome; (2) the continuous passage of the virus by its transient growth in cultures of T lymphocytes of normal blood donors; (3) the identification of a major protein associated with this virus, p25, not immunologically cross-reactive with the p24 of HTLV-I; and (4) the detection by immunoprecipitation of antibodies against this protein in two patients.

Essex and co-workers (1983)²⁷ detect antibodies cross-reactive with HTLV-I membrane protein in 25–30% of AIDS patients.

Gelman and co-workers (Gallo's group) (1983)²⁸ find evidence for presence of the viral genome of HTLV-I or an

from HTLVs; and (5) the antigenic cross-reactivity between core proteins of EIAV and LAV (virus isolates from LAS patients are called LAV, and from AIDS patients are called IDAV).

Gallo and co-workers (1984)³⁰ report the presence of HTLV-I antibodies in 10% of AIDS patients and isolates of HTLV-I or HTLV-II or variants of it in fewer than 10% of such cases.

March–April 1984. Montagnier *et al.* (1984)³¹ by using more sera of horses infected with EIAV confirm cross-reactivity of the core proteins of LAV with that of EIAV and identify a second viral protein, p18. Vilmer, Chermann, Montagnier and co-workers (1984)³² confirm previous isolation of an LAV-like virus from one haemophiliac and isolate another one

from his asymptomatic brother.

May 1984. Gallo's group (1984)³³⁻³⁶ reports: (1) mass and continuous production in a clone of a permanent cell line (H9) of HTLV-III from two AIDS patients and four additional isolates (SN, BK, CS, WT) also infectious for another clone (H4) derived from the same parental cell line (Popovic, Sarngadharan, Gallo and co-workers³³).

(2) 48 virus isolations, that is, 18 of 21 patients with pre-AIDS, 3 of 4 clinically normal mothers of juveniles with AIDS, 26 of 72 juveniles and adults with AIDS, 1 of 22 healthy male homosexuals, and 0 of 115 heterosexual subjects (Gallo, Salahuddin, Popovic and colleagues³⁴). The use of anti-p24 hyperimmune sera proves that the 48 isolates belong to the same kind of virus;

(3) The introduction of the Western blot technique for clinical detection of antibodies in 88% of 48 patients with AIDS, 79% of 14 homosexuals with pre-AIDS and less than 1% of hundreds of heterosexuals. A gp41 is identified as a major viral antigen (Sarngadharan, Popovic, Gallo and co-workers³⁵). Later, it is demonstrated by Veronese, Sarngadharan and Gallo to be the HTLV-III viral transmembrane component of the envelope⁶³.

(4) Partial characterization of the immunologically reactive proteins by the Western blot technique. (Schupbach, Sarngadharan, Popovic, Gallo and co-workers³⁶).

June 1984. Safai, Gallo, Popovic and Sarngadharan report 34 of 34 (100%) of AIDS patients positive for HTLV-III antibodies, 16 of 19 (84%) of LAS patients and 0 of 14 (0%) of controls (1984)³⁷.

Brun-Vezinet, Barre-Sinoussi, Montagnier, Chermann and co-workers (1984)³⁸ publish detection of antibodies against LAV proteins by ELISA in 74.5% of the patients presenting with lymphadenopathy syndrome, 37.5% of patients with frank AIDS, 18% of healthy homosexuals, 1% of blood donors.

July 1984. Kalyanaraman, Montagnier, Francis and co-workers (1984)⁴⁰ report the detection of anti-p25 (LAV) antibodies in 51 of 125 (41%) of AIDS patients, 81 of 113 (72%) of LAS patients, and 0 of 70 of healthy individuals; Montagnier and co-workers (1984)⁴¹ report the growth of LAV in continuous B-cell lines, most of them transformed by Epstein-Barr virus.

Klatzmann, Gluckman, Chermann, Montagnier and co-workers (1984)⁴² publish: (1) the selective isolation of LAV from CD4⁺ (T4⁺) lymphocytes of a healthy carrier of the virus; (2) the inhibition of CD4 cell growth at the same time of *in vitro* virus production; (3) the simultaneous disappearance of the CD4 antigen at the surface of the infected CD4 lymphocytes.

August 1984. A third group, Levy *et al.*,

(1984)⁴³ isolate virus antigenically and structurally related to LAV from San Francisco AIDS patients.

September 1984. Cheinsong-Popov, Weiss and collaborators publish identical prevalence of antibodies against antigens of HTLV-III grown in H9 line and of LAV-1 grown in CEM line in UK patients with AIDS or at risk of AIDS (1984)⁴⁴.

Goedert, Gallo and co-workers (1984)⁴⁵ report that in a cohort of homosexual men at risk of AIDS, 53% were antibody-positive for HTLV-III. In HTLV-III antibody positive subjects, AIDS developed at 6.9% per year.

October 1984. Brun-Vezinet, Montagnier, Piot, Quinn and co-workers (1984)⁴⁶ publish the presence of antibody against LAV in 35 of 37 Zairean patients with AIDS.

Zagury, Gallo and co-workers (1984)⁴⁷ isolate HTLV-III from cells cultured from semen of two patients with AIDS.

November 1984. Hahn, Gallo and co-workers (1984)⁴⁸ report the molecular cloning of HTLV-III virus.

Kitchen, Allan, Essex and co-workers (1984)⁴⁹, (1985)⁵⁰ identify the viral external glycoprotein gp120, a finding confirmed by Montagnier and co-workers (1985)⁵¹.

December 1984. Alizon, Barre-Sinoussi, Wain-Hobson, Montagnier and co-workers (1984)⁵² report the molecular cloning of LAV-1.

Wong-Staal, Shaw, Gallo and co-

workers (1984)⁵³ discover genomic heterogeneity of HTLV-III.

Dagleish, Weiss and co-workers (1984)⁵⁴ and independently Klatzmann, Gluckman, Montagnier and co-workers (1984)⁵⁵ show the CD4 molecule is involved in the receptor to the virus.

Popovic, Read-Connole and Gallo (1984)⁵⁶ publish a series of CD4-positive human neoplastic cell lines susceptible to and permissive for HTLV-III, including HUT78, Molt3 and CEM cell lines.

January 1985. The nucleotide sequence of the AIDS virus genome is established independently at the Pasteur Institute (1985)⁵⁷, at the NCI/NIH (1985)⁵⁸, at Genentech, Inc. (1985)⁵⁹ and at Chiron (1985)⁶⁰, revealing the similarity of the various isolates.

Sodroski, Wong-Staal, Gallo, Haseltine and co-workers (1985)⁶¹ demonstrate transactivation of transcription in HTLV-III infected cells.

Shaw, Gallo and co-workers (1985)⁶² discover the presence of virus in the brain.

March 1985. Redfield, Gallo and co-workers (1985)⁶⁴ describe heterosexual transmission of HTLV-III. □

We wish to acknowledge with gratitude the generous help of many of our colleagues in the preparation of this history. We cannot individually thank them all here, but would like to offer special thanks to Dr Jonas Salk for his help and guidance in completing this project.

1. Temin, H.M. & Mizutani, S. *Nature* **226**, 1211-1213 (1970).
2. Baltimore, D. *Nature* **226**, 1209-1211 (1970).
3. Hill, M. & Hillova, J. *C. r. Acad. Sci. Paris* **272D**, 3094 (1971).
4. Montagnier, L. & Vigier, P. *C. r. Acad. Sci. Paris* **274D**, 1977-1980 (1972).
5. Svoboda, I. *et al. J. gen. Virol.* **21**, 47-55 (1973).
6. Spiegelman, S. *et al. Nature* **237**, 1029-1031 (1970).
7. Gallo, R.C. *Nature* **234**, 194-198 (1971).
8. Sarngadharan, M.G., Sarin, P.S., Reitz, M.S. & Gallo, R.C. *Nature new Biol.* **240**, 67-72 (1972).
9. Gerwin, B.I. & Hilestein, J.B. *Proc. natn. Acad. Sci. U.S.A.* **69**, 2599-2603 (1972).
10. Morgan, D.A., Ruscatti, F.W. & Gallo, R.C. *Science* **193**, 1007-1008 (1976).
11. Barre-Sinoussi, F. *et al. Ann. Microbiol. (Inst. Pasteur)* **130B**, 349-362 (1979).
12. Poesz, B.J. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 7415-7419 (1980).
13. Rho, H.M. *et al. Virology* **112**, 355-360 (1981); Reitz, M.S. Jr *et al. Proc. natn. Acad. Sci. U.S.A.* **78**, 1887-1891 (1981); Kalyanaraman, V.S. *et al. Nature* **294**, 271-273 (1981); Robert-Guroff, M. *et al. J. exp. Med.* **154**, 1957-1964 (1981).
14. Hinuma, Y. *et al. Proc. natn. Acad. Sci. U.S.A.* **78**, 6476-6480 (1981).
15. Yoshida, M., Miyoshi, I. & Hinuma, Y. *Proc. natn. Acad. Sci. U.S.A.* **79**, 2031-2035 (1983).
16. Popovic, M. *et al. Nature* **306**, 63-66 (1982).
17. Catovsky, D. *et al. Lancet* **i**, 639-642 (1982).
18. Blattner, W.A. *et al. Int. J. Cancer* **30**, 257-264 (1982).
19. Kalyanaraman, V.S. *et al. Science* **218**, 571-575 (1982).
20. Gottlieb, M.S. *et al. Morbid. Mortal. weekly Rep.* **30**, 250-252 (1981); Gottlieb, M.S. *et al. N. Engl. J. Med.* **305**, 1423-1431 (1981).
21. Friedman-Kien, S.M. *et al. Morbid. Mortal. weekly Rep.* **30**, 305-308 (1981).
22. Siegel, P.P. *et al. N. Engl. J. Med.* **306**, 1439-1444 (1981).
23. Masur, H. *et al. N. Engl. J. Med.* **305**, 1431-1438 (1981).
24. Mildvan, D. *et al. Morbid. Mortal. weekly Rep.* **31**, 249-251 (1982).
25. Centers for Disease Control Task Force on Kaposi's Sarcoma and Opportunistic Infections *N. Engl. J. Med.* **306**, 1248-1252 (1982).
26. Barre-Sinoussi, F. *et al. Science* **220**, 868-871 (1983).
27. Essex, M. *et al. Science* **220**, 859-862 (1983).
28. Gelman, E.P. *et al. Science* **220**, 862-865 (1983).
29. Montagnier, L. *et al. in Human T-cell Leukaemia*

Lymphoma Viruses (eds Gallo, R.C., Essex, M.E. & Gross, L.) 363-379 (Cold Spring Harbor Laboratory, New York, 1984).

30. Robert-Guroff, M. *et al. in Human T-Cell Leukaemia Lymphoma Viruses* (eds Gallo, R.C., Essex, M. & Gross, L.) 288-296 (Cold Spring Harbor Laboratory, New York, 1984).
31. Montagnier, L. *et al. Ann. Virol. (Inst. Pasteur)* **186E**, 119-134 (1984).
32. Vilmer, E. *et al. Lancet* **i** 753-757 (1984).
33. Popovic, M., Sarngadharan, M.Q., Read, E. & Gallo, R.C. *Science* **224**, 497-500 (1984).
34. Gallo, R.C. *et al. Science* **224**, 500-503 (1984).
35. Sarngadharan, M.G., Popovic, M., Bruch, L., Schupbach, J. & Gallo, R.C. *Science* **224**, 506-508 (1984).
36. Schupbach, J. *et al. Science* **224**, 503-505 (1984).
37. Safat, B. *et al. Lancet* **i**, 1435-1440 (1984).
38. Brun-Vezinet, F. *et al. Lancet* **i** 1253-1256 (1984).
39. Deleted.
40. Kalyanaraman, V.S. *et al. Science* **225**, 321-323 (1984).
41. Montagnier, L. *Science* **226**, 63-66 (1984).
42. Klatzmann, D. *et al. Science* **225**, 59-69 (1984).
43. Levy, J.A. *et al. Science* **225**, 840-842 (1984).
44. Cheinsong-Popov, R. *et al. Lancet* **ii**, 477-480 (1984).
45. Goedert, J.J. *et al. Lancet* **ii**, 711-716 (1984).
46. Brun-Vezinet, F. *et al. Science* **226**, 453-456 (1984).
47. Zagury, D. *et al. Science* **226**, 449-451 (1984).
48. Hahn, B.H. *et al. Nature* **312**, 166-169 (1984).
49. Kitchen, L.W. *et al. Nature* **312**, 367-369 (1984).
50. Allan, J. *et al. Science* **226**, 1091-1094 (1985).
51. Montagnier, L. *et al. Virology* **144**, 283-289 (1985).
52. Alizon, M. *et al. Nature* **312**, 757-760 (1984).
53. Wong-Staal, F. *et al. Science* **227**, 759-762 (1985).
54. Dagleish, A.G. *et al. Nature* **312**, 763-767 (1984).
55. Klatzmann, D. *et al. Nature* **312**, 767-768 (1984).
56. Popovic, M., Read-Connole, E. & Gallo, R.C. *Lancet* **ii**, 1472-1473 (1984).
57. Wain-Hobson, S., Sonigo, P., Danos, O., Cole, S. & Alizon, M. *Cell* **40**, 9-17 (1985).
58. Ratner, L., Gallo, R.C. & Wong-Staal, F. *Nature* **313**, 636-637 (1985).
59. Hausing, M.A. *et al. Nature* **313**, 460-468 (1985).
60. Sanchez-Pescador, R. *et al. Science* **227**, 484-492 (1985).
61. Sodroski, J. *et al. Science* **227**, 171-173 (1985).
62. Shaw, G.M. *et al. Science* **227**, 177-182 (1985).
63. Veronese, F. *et al. Science* **229**, 1402-1405 (1985).
64. Redfield, R.R. *et al. J. Am. med. Ass.* **253**, 1571-1573 (1985).

Restoring technological competitiveness

At the first of series of meetings organized by Nature the Labour party spokesman on science and technology, Dr Jeremy Bray MP, presented his party's policies on science and technology. Five interlocutors and an audience of eighty provided questions.

THE Labour party is considering a radical scheme of financial incentives that could lead to an increase of £4,000 million a year in expenditure on civil research and development. Its aim is to help reduce the widening gap between research expenditure in Britain and other industrialized countries, which the party believes is the cause of Britain's inability to compete successfully in world high-technology markets. The scheme would be accompanied by a restructuring of defence research, a renewal of the science base and changes in the machinery of government to give science a stronger voice in national affairs. Finance for the scheme would be provided by an increase in corporation tax.

These were the key points by Dr Jeremy Bray, Member of Parliament for Motherwell South, a mathematician and Labour party spokesman for science and technology, in a draft white paper (policy document) prepared for the first *Nature* meeting on the political parties' plans for science*.

The Labour party is already committed, in its official policy document *New Industrial Strength for Britain — Labour Programme for National Renewal*** to mounting "a major drive to increase research and development including measures to assist industry and universities". The statement says that Labour would increase expenditure on the universities, appoint a Minister of Science and Technology and establish both a Council of Science and Technology and an Office of Technology Assessment. Bray's draft white paper attempts to provide a concrete plan for action to stimulate research and development.

The details, if approved by ministers in office, could "make it possible for a new government in the early days of office to issue a white paper on research and development to set that action in train".

Britain's decline

The picture painted of Britain's decline in the draft white paper is familiar. Its assumption, accepted by all political

parties, is that the advanced nations can retain their high standards of living only by continuously introducing new technology and new products. Branch factories, using already established technology, can



Dr Jeremy Bray MP aims to boost research and development expenditure.

be run, Bray points out, as efficiently in Korea, Taiwan, Hong Kong and Singapore as they can in Britain, but at a fraction of the labour cost. Behind these newly industrialized countries come a host of others that will soon be able to offer the skills efficiently to run advanced production processes.

New technology requires massive investment in research and development. All the evidence is that Britain is failing to make that investment. On the most basic measure of all, research and development expenditure relative to gross domestic product (GDP), Britain alone of the industrialized nations is failing to move forward (see the figure). Relative expenditure has been static for two decades during which both West Germany and Japan have increased their spending by 1.3 per cent of GDP. But it is not simply a ques-

tion of the totals: Britain's pattern of expenditure is unique.

Among five industrialized countries (the United Kingdom, West Germany, Japan, the United States and France) the United Kingdom's expenditure on defence research is the highest in relation to expenditure on civil research (46 per cent against West Germany's 4 per cent and Japan's 0.5 per cent). Furthermore, the percentage of civil research and development funded by industry, rather than government, is lower in the United Kingdom than in any country other than France (United Kingdom 53 per cent, West Germany 60 per cent, Japan 80 per cent, France 35 per cent). The two statistics taken together paint a dismal picture of British industry dependent on government defence research funds and failing to invest in its own civil technology. That, in Dr Bray's view, goes a long way to explaining Britain's inability to provide high-technology consumer goods. Overall, the output of 'high research intensity goods' has increased in Britain, but at an average rate of only 3 per cent, well below world levels. Imports as a percentage of sales have almost doubled in the past ten years. The cure, in Bray's view, is to persuade industry to invest more heavily in research and development.

Incentive scheme

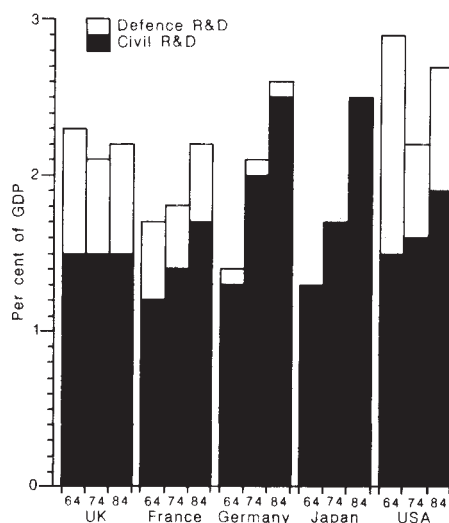
To make industry spend more, Bray proposes that government should provide a large percentage of any increases made in individual companies' research and development expenditure, subject to upper limits set by the number of employees in a company. The initial proposal (put forward for "discussion and refinement") is that grants of 60 per cent of the increase over 1986-87 levels be given.

Most other countries have incentive schemes already in place. At 60 per cent, Bray's scheme would be highly geared, but still not equal to Australia's new scheme where a 150 per cent allowance of total (rather than the increase in) research and development expenditure is granted. What effect would this have on expenditure and the economy?

According to Bray's calculations at the end of four years civil research and development expenditure would increase by £4,000 million a year from its present level of £4,950 million, mostly in industry. That, he claims, would lead over a longer

*This meeting was the first of three at which representatives from the three main British political parties will put forward their policies for science. Details of later meetings will be announced nearer the time. The Labour party draft white paper is available from Dr Jeremy Bray at the House of Commons, who would welcome comments on it.

** Available from the Labour Party, 150 Walworth Road, London SE17 1JT, price £1.



Research and development expenditure as a percentage of gross domestic product.

period, to a call for an increase of fixed investment of £15,000 million per annum, and of manufacturing and marketing start up costs of a further £15,000 million per annum. An increase in sales of £60,000 million and direct employment of 1 million would be generated.

Alongside these changes, the science budget (that is, the budgets of the University Grants Committee and the research councils) would be increased by £400 million a year to take it to £1,500 million a year. Cuts totalling £1,100 million would come in the £2,390 million a year defence research and development budget, on the grounds that more efficient use of funds could be made through increasing co-operative projects with other European nations, perhaps coordinated by the North Atlantic Treaty Organization.

Defence cuts alone would not pay for the incentive scheme. The bulk of the money would be raised through an increase in corporation tax (perhaps from 35 per cent to 40 per cent). The increase in tax, paid by the whole of industry, would be immediately recycled to manufacturing industry in research and development grants at no cost to the government.

Basic research

The case for strengthening basic research is "put in terms of its necessary contribution to the recovery of national technological competitiveness and the consequent economic benefits". The research councils would continue "in their present pattern" But Bray makes a more controversial suggestion that the renewal of Britain's science base could be aided by inviting help from the international community of scientists to set an "agenda for science". He suggests that the Royal Society and the research councils could organize international meetings at government expense to set that agenda.

Manpower

The expansion of research and development cannot be carried out if more scientists are not found. Some defence scientists would be redeployed. Other steps would be taken to make it easier for researchers to move between industry, research establishments, higher education and schools, and suitable refresher courses would be provided. One possibility Bray is considering is a "research service" that would take an interest in the career of a scientist "wherever he or she worked". But in the initial period of high growth it would be necessary to tempt engineers and scientists from India and from among the overseas Chinese to Britain to fill the gaps.

Machinery of government

The new programme is on such a scale and involves such a wide range of government departments that a new post of Minister of Science and Technology would have to be created to coordinate overall efforts. Actual policy implementation would, however, remain with individual departments. Direct executive power would be given to the minister only in the management of the research councils, manpower supply, and the running of an information service.

A Council for Science and Technology, chaired by the prime minister, would be set up as the government's central consultative body on science and technology and would replace the present Advisory Council for Applied Research and Development. The idea is that it should spread its roots wide: into "every department and boardroom and laboratory in the country". And, on top of these changes and most important of all, in Bray's closing words, would be "a determination on the part of the scientific community to present itself as an integral part of our national industrial, economic, political, and cultural scene so it is not simply a declining outpost of past national glory".

Will industry respond?

The heart of Labour's policy is that subsidizing research and development and expanding its scale will provide the key to Britain's economic recovery. But are there not more intractable reasons for Britain's decline than a failure to spend enough on research? After all, if industry could increase profits through increasing research expenditure why does it not do so? Bray believes that it is because of the "need for managements to maximize short-term profitability to satisfy short-sighted investors and avoid take overs". The first of the interlocutors raised some doubts.

Dr Geoffrey Cooper (director, Technical Change Centre) pointed out that companies that already value research (Glaxo, ICI, Unilever and so on) will be

little affected by the incentive scheme. Elsewhere, the real difficulty may not be lack of money or patience but in making everyone within a company aware of research and development. The gulf between the scientist and the practical businessman remains wide. Professor Martin Rees (University of Cambridge) took up the point, stressing that it is hard to see how companies now spending tiny amounts on research would make effective use of increased funding. "Could it not be that what one needs in most of industry is simply a more educated management and workforce?" And Professor Denis Noble (University of Oxford) asked whether efforts should not be made to put scientists onto the boards of companies. Bray agreed that there is insufficient awareness in much of industry to make good use of the incentive scheme. But he is optimistic, believing that it would be possible to set up "research clubs" in which advanced companies would provide leadership and help spend research money well. Subcontracting of research is provided for in the scheme.

Intellectual plunder?

Peter Rigby (National Institute for Medical Research) was clearly not alone in wondering where the manpower for such a large-scale increase in research would come from. And he clearly had supporters when he stressed that replacing Britons who had joined the brain-drain by "plundering the intellectual resources of the third world would be quite improper". But Bray saw it differently, giving those in the third world a "chance to pinch our ideas. They could work in leading edge teams and go back to establish their own far faster than if there was no international mobility". Neil Kinnock, leader of the Labour party, has already received a warm welcome for the idea from directors of Indian research institutes in New Delhi, he said.

Unpredictable change

The agenda for science also came under fire. Professor Sir Hans Kornberg (University of Cambridge) summed up criticism by saying that if "we look back ten years and see what we would have predicted as growing points that would change the world we get practically no answers right". It is better to "assume that good people will be in touch with good scientists and know where the exciting opportunities are". But others felt some sort of agenda, like that appearing in Japan's Human Frontiers Science Program could help restore morale. The collapse of self-confidence among British scientists was clearly felt acutely by the audience, particularly by those on short-term contracts. Bray hoped that a "research service might help" but "a lot of working out of the idea is needed".

Superconductivity by rabbits' hats

Nobody quite understands why there is such excitement about the succession of superconducting materials with ever-higher transition temperature. But the phenomenon may be simpler than it seems.

THE first thing to say about the new superconducting materials based on copper oxide is that they have nothing in common with any known form of that material (cuprous or cupric, as the case may be). Moreover, the rare-earth oxide solid solutions with copper oxide are not homogeneous mixtures, but layer structures. That explains why the current-carrying capacity of these materials is limited; unlike more familiar materials such as niobium, or lead, the superconducting state is probably not annihilated with increasing voltage by the self-interaction of its own magnetic field, but by a physical breakdown of the material.

Here is a simple model of such a superconductor. Take a vessel filled with anhydrous benzene, the molecules of which are supposed to be held together with extra binding energy by the mutual interaction of electrons centred on neighbouring carbon atoms. But carbon atoms have four mutually interchangeable electrons. (Two, with lesser energy, are interchangeable with each other and form the core of the electronic shell.) Of the several ways of enumerating the states of the four outer electrons, mere symmetry implies that the most appropriate must be that in which three of the four second energy-level electrons are in quantum mechanical orbits of asymmetrical dumbbell-shape making an angle of 120° ($2\pi/3$) with each other. Symmetry (or Schrödinger's wave equation) then requires that the fourth electron should be at right angles to the other three and should have the shape of a symmetrical dumbbell. The extra stability of the benzene molecule arises from the mutual interaction of electrons in these fourth orbits. Linked together on the scale of a whole molecule, they are known to molecular chemists (and to many people still at high-school) as π -orbitals.

So what happens if you take a quantity of benzene, somehow arrange that all the six-sided molecules are oriented with their faces parallel to each other and then apply a changing magnetic field in the perpendicular direction? Superficially, one would guess that the 'delocalized' electrons in the π -orbitals are free to move around the benzene rings as if they were free electrons in a metal.

The snag is that Pauli's exclusion principle supervenes, making sure that if there are three electrons moving in one direction around the benzene ring, there must be three others (nobody can hope to tell

which they are) on the average moving in the opposite direction. That, for what it is worth, is why graphite, a two-dimensional array of benzene rings, is not a metal but a semiconductor. (The conductivity increases with temperature.) At lowish energy (or temperature), Pauli's principle will require that there should be equal numbers of electrons moving in one direction and its opposite; as the temperature is increased, the chance of promoting such an electron into an unpaired state of mobility will be increased. (The 'hole' left in the valance band will also be mobile, and in the opposite direction, but will appear more massive so less mobile.)

So could graphite, or even benzene, be made superconducting? Such a phenomenon plainly requires some kind of cooperative interaction between moving electrons and the vibrations of the underlying lattice. Moreover, it is not sufficient that some forms of electronic motion should be favoured (or selected) by some forms of lattice vibration, but that the opposite should also be true. Superconducting (or cooperatively diamagnetic) benzene might be found in the nanokelvin region. Superconducting graphite layers are topologically even less accessible. But the feasibility of making semiconducting and even superconducting linear polymers has been much canvassed in the past few years. There is a great deal to be said for it.

So this is how a copper-oxide compound functions as a superconductor. First, after the model of lanthanum copper oxide (like graphite, a semiconductor at all known temperatures), these materials are layered structures in which sheets of copper (and oxygen) atoms alternate with sheets of rare-earth (and oxygen) atoms. What the participants in the race for a room-temperature superconductor have found is that success hangs on the clever manipulation of two parameters — the extent to which lanthanum atoms are replaced by other rare-earth atoms or even mundane materials such as calcium, barium and strontium, and the extent to which the material as a whole is stoichiometrically deficient in oxygen. What that suggests is that the alternative planes containing copper atoms are those in which superconduction happens, and that the planes carrying the rare-earth (or mere group II) function as extraplanar means of adjusting the electrical potentials and thus the energy levels of the conduction bands, not to mention the effective

masses of the oppositely travelling holes that may be created in them either by the promotion of electrons into the conduction bands or by a general deficiency of electrons in the rare-earth sheets.

In that sense, the rare-earth layers function as do the grids in an old-fashioned thermionic valve, or tube, by facilitating (or otherwise) a conduction process. But in the case of the ternary copper oxides, the trick is so to bias the parameters of electrical conduction that there is a reasonable chance of finding a cooperative match between the motion of an electron (or of a 'heavy' hole) and a lattice motion. Ideally, there should be a range of frequency over which the motions overlap. The best bets for those who seek a room-temperature superconductor are plainly those in which the alternating layers of the copper and rare-earth oxides span as great a range of vibration frequencies as possible, but questions of electronegativity and electron-band density (which determine effective masses of electron holes) also matter.

So the search for even better superconductors will be in the hands of the structural solid-state people. Analogues of lanthanum copper oxide will be the favourites for a time. Schemes for packing the spaces between graphite layers in such a way as to make the whole structure are probably unfeasible, given the low vibration frequencies they would generate, but the notion that it may be possible to design superconductors for particular purposes, and then to synthesize them, perhaps by the techniques of epitaxial growth now widely used in semiconductor electronics, are not out of court.

There are several obvious caveats to be made in such a simple tale. Those who say that this is a crude way of describing the Frolich picture of superconductivity should think again; electrons and holes cannot move without reference to each other. It may also be a better picture to say that the electrons move in one kind of layer and the holes in the other, at least on the average. The fact is that the demonstration of one kind of two-dimensional superconductivity at high temperature relative to absolute zero should make it possible in due course to find materials superconducting at red heat. But true bulk superconductors, requiring biasing electronegativities everywhere, may be topologically impossible.

John Maddox

Australian views of the supernova

AUSTRALIA has the only major radio telescopes in the world in a position to observe the supernova SN 1987a, the main instruments being the CSIRO Parkes radio telescope and the Sydney University Molongo Synthesis and Fleurs radio telescopes. All saw a radio flash corresponding to the optical flash of SN 1987a within the first few days of the explosion, but after a week the flash had faded so that it could be seen only by the sensitive 1.6-km-long array of the Molongo Synthesis telescope. Now all the radio-astronomers are waiting for the expanding shell of gases to grow large and transparent enough to release more of the vast amounts of bound energy in the radio part of the spectrum, which could take up to 100 days. The initial optical/radio flash is an entirely new phenomenon.

When strong radio emission is again detected radioastronomers hope to have in place Australia's first very long baseline interferometer. They are now working furiously to link the Parkes telescope, a dish belonging to the US deep-space tracking network near Canberra, a 26-metre dish at the University of Tasmania and a satellite tracking dish near Alice Springs. Following the supernova should keep them busy for the next 10 years. They will be looking hard for any sign of a newborn pulsar at the supernova's heart.

The optical astronomers are at the Anglo-Australian Observatory (AAO), which has a 3.9-metre telescope, the AAT, and at the Mount Stromlo and Siding Spring observatories (MSSSO) which have 1.9- and 2.3-metre telescopes. The AAO is excising an hour from each night's allocation to observe SN 1987a with whatever instrument is attached to the telescope at the time. The MSSSO, on the other hand, is making regular nightly observations for a full record of the supernova's evolution.

Both these observatories are monitoring the speeds of the expanding shells of gas by observing Doppler shifts in high-resolution spectra. The peak was $25,000 \text{ km s}^{-1}$ on 27 February. The velocity has been dropping ever since but that does not mean that the shells are slowing down, rather that the fastest outermost layers are becoming transparent so that deeper, slower-moving levels can be seen. Narrow-band polarization measurements made with the AAT show that the supernova is not symmetrical.

Both optical observatories will attempt speckle interferometry to measure the angular size of the supernova, a vital observation if type II supernova are to be used as standards to measure the distance to distant galaxies. The AAT has the highest resolution, about 20 mas. But when the first observations are made on 2 April it is unlikely that SN 1987a will be bigger than 10 mas.

Charles Morgan

Evolution

Sex determination in fish

Michael Bulmer

IN most animals the sex of an individual is fixed at conception by its genotype, but in some species it is determined by environmental conditions after conception. In many turtles, for example, sex is determined by the incubation temperature of the eggs, with males being produced in cool and females in warm conditions¹. Light is thrown on the adaptive significance of this phenomenon by D.O. Conover's study, reported on page 496 of this issue², of geographical variation in the Atlantic silverside, a fish with both genotypic and temperature-dependent sex determination.

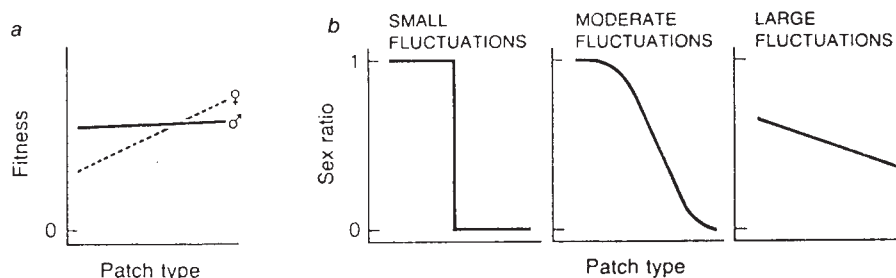
It has been suggested³ that environmental sex determination is favoured when environmental effects have different consequences for the two sexes. Suppose that some juveniles encounter good conditions and grow into large adults, while others are less fortunate and grow into small adults. Suppose also that adult size matters more to a female than to a male because female fecundity is directly proportional to size whereas male mating ability is little affected by size (*a* in the figure). Then it would be advantageous for a juvenile to become a male or a female depending on whether it finds itself in a poor or a good environment (*b* in the figure) because the reward for being large is greater for a female whereas the penalty for being small is less for a male.

How can such an adaptive response be brought about? In haplodiploid animals, such as ants, bees and wasps, the mother can control the sex of an offspring by whether or not she fertilizes the egg, unfertilized eggs becoming males and fertilized eggs females. This gives the mother the ability to vary the sex of her offspring adaptively. Parasitic wasps often lay a male egg in a small host and a female egg in a large host, in accordance with this theory⁴.

It is much more difficult for animals with the usual chromosomal mechanism of sex determination to vary the sex of

their offspring adaptively, and there is thus selection to evolve a mechanism whereby sex is determined directly by the environment of the offspring. Conover has extensively studied the Atlantic silverside^{2,5-7}, and he finds that most breeding fish are one year old and spawn *en masse* at fortnightly intervals (new and full Moon) in May, June and July off the eastern seaboard of the United States. Young fish grow throughout their first summer, those hatching earlier growing to a larger adult size than those that hatch later. Female fecundity is more strongly related to size than is male fecundity. Temperature has a strong influence on sex determination, the sex ratio being female-biased at low and male-biased at high temperatures. This can be regarded as an adaptive response to the relationship between hatching date and body size, with sea temperature at the critical age for sex determination being used as an environmental cue to hatching date.

The danger of this strategy is that it may produce a large excess of males one year and an excess of females another year if there are large annual fluctuations in temperature, whereas an approximate equality of the two sexes is selectively advantageous (the optimal sex ratio is slightly male-biased⁸). In this situation, theoretical calculations⁸ show that the optimal strategy is a mixture of environmental and genotypic sex determination, producing a graded response buffering the fluctuations. This is just what the silverside does. Genetic crosses show that sex is determined by a complex interaction of temperature with major sex-determining loci^{6,7}. The geographical comparison described by Conover in this issue² of populations from South Carolina to Nova Scotia shows that the strength of the genetic compared with the environmental component of sex determination increases with latitude; indeed, in Nova Scotia there is no temperature dependence, the sex ratio being 0.5 at all temperatures. This



The effect of environmental fluctuations on the selected sex-ratio¹. When fitness varies differentially with patch type (*a*), the optimal sex-ratio response (*b*) is a step function when there are only small fluctuations in patch-type frequencies, but becomes attenuated as they increase.

effect can be attributed to the decrease in the length of the breeding season with increasing latitude. A shorter breeding season both lessens the advantage of early over late breeders and magnifies the variability in mean temperature during the breeding season, thus selecting for an increase in genotypic over environmental sex determination.

A similar system has been reported in the shrimp *Gammarus duebeni* (ref. 9 and C. Naylor, J. Adams and P. Greenwood, in preparation). The main difference here is that adult size matters more to males than to females because males can only pair with females smaller than themselves. In northern latitudes there is one generation per year and sex-determination responds to day length, with males being produced at short day lengths at the beginning of the season and females at long day lengths at the end of the season. This process is adaptive because adult males will have more opportunity to grow larger than adult females when they breed the following year. Further south, there are two overlapping generations per year and the response of sex determination to day length, which would no longer

be adaptive, disappears.

These examples make good evolutionary sense, but the evolution of temperature-dependent sex determination in reptiles remains enigmatic^{1,10}. In many turtles males are produced in cool and females in warm conditions, with an abrupt transition from all-male to all-female broods; in alligators, crocodiles and some lizards the opposite is true (females in cool and males in warm conditions), whereas other lizards have genotypic sex determination. There is still no convincing adaptive explanation for these striking facts. □

1. Bull, J. J. *Evolution of Sex Determining Mechanisms* (Benjamin/Cummings, Menlo Park, 1983).
2. Conover, D. O. & Heins, S. W. *Nature* **325**, 496-498 (1987).
3. Charnov, E. L. & Bull, J. J. *Nature* **266**, 828-830 (1977).
4. Charnov, E. L. *The Theory of Sex Allocation* (Princeton Univ. Press, 1982).
5. Conover, D. O. *Am. Nat.* **123**, 297-313 (1984).
6. Conover, D. O. & Kynard, B. E. *Science* **213**, 577-579 (1981).
7. Conover, D. O. & Heins, S. W. *Copeia* (in the press).
8. Bulmer, M. G. & Bull, J. J. *Evolution* **36**, 13-26 (1982).
9. Adams, J. & Greenwood, P. J. *J. Zool.* **211**, 35-46 (1987).
10. Hutton, J. M. *J. Zool.* **211**, 143-155 (1987).

Michael Bulmer is in the Department of Biomathematics, University of Oxford, 5 South Parks Road, Oxford OX1 3UB, UK.

DNA replication

Many strands converge

Julian Blow

MANY areas of research in molecular biology are concerned with the factors that drive the cell through its division cycle. Three papers in this issue¹⁻³ unexpectedly draw together work on viral DNA replication, cell-cycle-regulated protein synthesis and the cellular DNA polymerases. The first paper¹ shows the requirement of an *in vitro* DNA replication system for proliferating cell nuclear antigen (PCNA, also called cyclin), a protein synthesized only during the S phase of the cell cycle⁴. This result suggests a direct role for PCNA in DNA replication. The additional discovery of identity between PCNA and a polymerase- δ auxiliary factor^{2,3} raises questions about the widespread acceptance of polymerase- α as the sole polymerase required for eukaryotic DNA replication.

PCNA already has an interesting history. It is recognized by abundant antibodies in certain patients suffering from autoimmune disease, and is present only in the nuclei of proliferating cells. Several years ago it was reported⁵ that PCNA is identical to cyclin. Therefore, PCNA synthesis and nuclear localization occur only during the period of DNA synthesis (see figure). Because the name 'cyclin' has been given to at least two different proteins, the name PCNA is preferable to avoid confusion. Subsequently its syn-

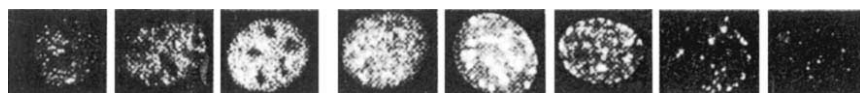
thesis and distribution was shown to be dependent on progression through S phase, suggesting that it is involved in DNA replication⁶. This prediction has been amply fulfilled.

In the paper on page 471 of this issue¹,

stage of replication more than fivefold. Perhaps a requirement for PCNA is one of the reasons why SV40 induces its host cell to replicate.

The two other papers in this issue, by Bravo *et al.*² on page 515 and Prelich *et al.*³ on page 517, provide further indications of the role of PCNA. Both show the identity of this protein with a protein previously purified by its ability to stimulate the activity of calf thymus DNA polymerase- δ up to 1,000-fold on certain templates⁸. The protein has no such stimulatory effect on DNA polymerase- α . There is no direct evidence that the role of PCNA in either SV40 or cellular replication is related to its effect on polymerase- δ . But its implication in the chain-elongation stage of SV40 replication *in vitro*, and its implication in DNA synthesis *in vivo*, make this an attractive possibility. The roles of the two polymerases in eukaryotic replication are therefore called into question.

DNA polymerase- α has been accepted as the sole polymerase involved in eukaryotic DNA replication. Much of the evidence for this is based on studies of inhibitors of polymerases α , β and γ , which eliminate the necessity for either β or γ in replication. There are four main lines of evidence for the involvement of polymerase- α : first, all its inhibitors inhibit replication; second, mutants temperature-sensitive in polymerase- α are also temperature-sensitive in replication; third, on microinjection, antibodies against polymerase- α inhibit replication; and finally, its activity rises in cells synthesizing DNA. None of this evidence excludes the possibility of another polymerase being required in addition to α . One objection to polymerase- α as the replication



G₁/S ————— S phase —————> S/G₂

Progression of the nuclear distribution of PCNA during the S phase of 3T3 cells. Cells stained with antibodies against PCNA show nuclear fluorescence only during S phase (the period of DNA synthesis). This is also the only period when PCNA is synthesized. (From ref. 6.)

Prelich *et al.* identify a factor required for efficient simian virus 40 (SV40) DNA replication *in vitro*. There is interest in the replication of SV40 viral DNA because it requires, apart from one viral protein, only cellular components, and may therefore be a good model for cellular replication⁷. Using a cell-free system from human 293 cells that replicates DNA containing the SV40 origin of replication in the presence of T antigen, Prelich *et al.* show that PCNA is one of the cellular proteins required for efficient viral DNA replication. They identify it as a factor that increases the efficiency of the chain-elongation

enzyme⁹ is its lack of 3'-5' proofreading exonuclease activity, which can remove a mismatched nucleotide directly after it has been added. Without this activity there are problems in distinguishing which member of a mismatched pair is in error once polymerization has progressed past the mismatch.

In 1976, Byrnes *et al.* isolated⁹ a new eukaryotic DNA polymerase from rabbit bone marrow, which they called δ . They distinguished δ from α by four criteria: first, possession of a 3'-5' exonuclease activity; second, a distinct profile of polymerase activity on different templates

(unaffected by inhibition of the exonuclease activity); third, a different relative molecular mass; and finally, a different purification procedure. They suggested that α and δ are produced by different modifications of a common precursor. When a new polymerase was subsequently isolated from calf thymus¹⁰ it was also designated δ on the grounds that it possesses a 3'-5' exonuclease activity and that its template specificity resembles (but is not identical to) that of the rabbit polymerase- δ . However, polymerase- δ from the two sources have different molecular masses and subunit composition.

The relationship between polymerases α and δ remains uncertain. They have very similar sensitivities to inhibitors, including aphidicolin which is often used as a diagnostic for the involvement of polymerase- α . The nucleotide analogues BuPdGTP and BuAdATP seem to be the only inhibitors capable of making a distinction between the polymerases, with α sensitive and δ resistant¹². As purification procedures have improved, high-molecular-mass forms of polymerase- α have increased in size, one associated with 3'-5' exonuclease activity¹³. Similarly, high-molecular-mass forms of polymerase- δ have been reported with DNA primase activity (like α) and moderate sensitivities to BuPdGTP and anti-polymerase- α antibodies¹¹. Much work remains to be done to determine the relationships between the growing number of high-molecular-mass forms of these polymerases.

Are there two distinct forms of replicative polymerase within the living cell? If so, the requirement for PCNA in the elongation stage of SV40 replication implies that polymerase- δ is essential for this step. Polymerase- α has potential roles in the initiation/priming stage (as suggested by Prelich *et al.*¹) and/or in chain elongation in conjunction with polymerase- δ . Leading and lagging strand synthesis at the replication fork are functionally distinct processes. One possibility, requiring both polymerases for chain elongation, is that each polymerase is specialized either for continuous or for discontinuous synthesis at a fork. The tight association of DNA primase activity with polymerase- α might favour a role for this enzyme on the lagging strand. Alternatively, the difference between polymerases α and δ could merely reflect differences in isolation procedures from a common precursor *in vivo*. Proteolysis, subunit loss or other similar modification during purification may have significant effects on enzymatic behaviour. It might be relevant that the two polymerases from rabbit and calf thymus are expected to respond differently to PCNA. Calf thymus polymerase- δ uses poly(dA)/oligo(dT) as a template only inefficiently, when it is strongly enhanced by PCNA⁴, but rabbit polymerase- δ alone is active on this template¹².

Whatever the conclusion concerning polymerase- δ , PCNA emerges as a very interesting protein. Most studies of eukaryotic DNA replication implicate initiation, rather than chain elongation, as the key controlling event. Why, then, is PCNA synthesis so tightly controlled if, as suggested by the work reported in this issue, it is involved in the elongation stage of replication? Is its role in the SV40 replication system related to its stimulation of DNA polymerase- δ ? Eukaryotic cell-free DNA replication systems should help solve these and other related problems in the complex tangle of the threads of cellular replication. □

Astronomy

Motions in the local Universe?

Carlos Frenk and John Lucey

IT IS more than 50 years since Edwin Hubble discovered that galaxies are receding away from us with a velocity approximately proportional to their distance. An understanding of the details of this flow is a requisite for understanding the structure of our Universe on large scales. Important new developments in this subject were discussed at a recent workshop hosted by the Royal Greenwich Observatory*.

The expansion discovered by Hubble represents an average motion. In general, 'peculiar motions', that is, departures from the mean flow caused, for example, by gravitational perturbation on galaxies or galaxy clusters, are expected. If, as most astronomers believe, our Universe is homogeneous in the mean, such distortions should average out as increasingly large regions are considered. The cosmic microwave background (CMB) provides a convenient rest frame against which motions can be referred. When measured from the Earth the CMB radiation exhibits a characteristic dipolar pattern which is thought to result from the 'peculiar' motion of our Local Group of galaxies moving with a speed of about 600 km s⁻¹. Attempts to verify the large-scale smoothness of the Universe by averaging the peculiar velocities of distant galaxies (or, equivalently, by determining the velocity of the Local Group relative to those galaxies) have produced much controversy.

The first indication of something odd came with the pioneering study¹ of Rubin *et al.*, whose measurements implied that a distant sample of 96 spiral galaxies appear to be streaming with a velocity of about 1,000 km s⁻¹ in the CMB frame. This result was regarded with scepticism by many, especially after Hart and Davies reported² that a sample of 84 somewhat less-distant spirals reproduced the rest frame defined

1. Prelich, G. *et al.* *Nature* **326**, 471-475 (1987).
2. Bravo, R. *et al.* *Nature* **326**, 515-517 (1987).
3. Prelich, G. *et al.* *Nature* **326**, 517-520 (1987).
4. Bravo, R. & Celis, J.E. *J. Cell Biol.* **84**, 795-802 (1980).
5. Mathews, M.B. *et al.* *Nature* **309**, 374-376 (1984).
6. Bravo, R. & Macdonald-Bravo, H. *EMBO J.* **4**, 655-661 (1985).
7. Li, J.J. & Kelly, T.J. *Proc. natn. Acad. Sci. U.S.A.* **81**, 6973-6977 (1984).
8. Tan, C.-K. *et al.* *J. biol. Chem.* **261**, 12310-12316 (1986).
9. Byrnes, J.J. *et al.* *Biochemistry* **23**, 1906-1913 (1984).
10. Lee, M.Y.W.T. *et al.* *Biochemistry* **23**, 1906-1913 (1984).
11. Crute, J.J. *et al.* *Biochemistry* **25**, 26-36 (1986).
12. Byrnes, J.J. *Biochem. biophys. Res. Commun.* **132**, 628-634 (1985).
13. Ottiger, H.-P. & Hubscher, U. *Proc. natn. Acad. Sci. U.S.A.* **81**, 3993-3997 (1984).

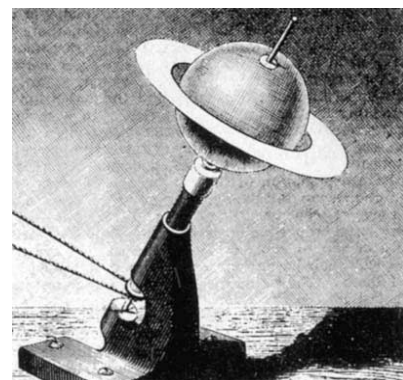
Julian Blow is in the Cancer Research Campaign Molecular Embryology research group, Department of Zoology, Downing Street, Cambridge CB2 3EJ, UK.

by the CMB. The impression of a largely quiescent Universe was reinforced³ by Aaronson *et al.* who found a sample of 10 distant galaxy clusters to have negligible random motions.

The issue returned to prominence early last year at a meeting held in Hawaii, discussed in a News and Views article⁴ by Bond and van den Bergh, where two new independent sets of measurements were presented. First, infrared observations⁵ by Collins *et al.* of 45 of the spirals studied by Rubin *et al.* in the distance range 36-65 h⁻¹ Mpc (where h is the Hubble constant in units of 100 km s⁻¹ Mpc⁻¹) confirmed the original result. Second, an all-sky survey⁶ of more than 400 elliptical galaxies within 60 h⁻¹ Mpc by Burstein *et al.* suggested that this local region is flowing coherently at about 700 km s⁻¹. It was immediately recognized that distortions in the Hubble flow of this size would probably conflict with nearly all currently popular theories⁷

100 years ago

AERIAL VORTICES AND REVOLVING SPHERES



THE guard of the revolving sphere is taken away, and we place parallel to its equator a circle of paper with an interior diameter greater than the exterior diameter of the sphere; the circle maintains itself in the plane of the equator.

From *Nature* **35**, 514; 31 March 1887.

*Workshop at the Royal Greenwich Observatory, Herstmonceux Castle, UK 12 - 13 January 1987.

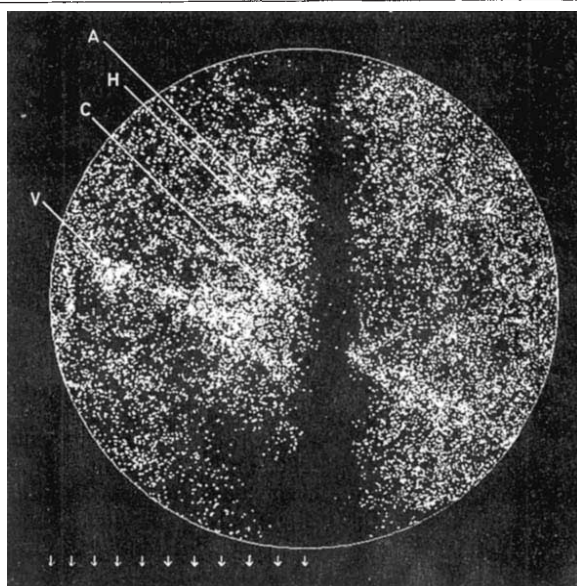
of the growth of cosmic structure. The new data presented at the Royal Greenwich Observatory meeting, together with a detailed reappraisal of previous data, produced a rather different viewpoint from the one that emerged at Hawaii. Although the controversy is by no means over, there is general agreement that streaming motions have been detected but that the flow pattern is considerably more complex than had hitherto been recognized. Early, far-reaching conclusions may have been premature.

R. Davies (Jodrell Bank) reported new H I observations of a sample of spirals; this enlarged sample, at an average depth of $30 h^{-1}$ Mpc, no longer reproduces the CMB rest frame. In a related study of an all-sky sample of spirals in the distance range $5-60 h^{-1}$ Mpc, L. Staveley-Smith (Jodrell Bank) finds evidence for streaming motions as well; both low- and high-redshift subsamples yield substantial velocities but the amplitudes and directions of the various solutions disagree with one another. Collins (University of Edinburgh) presented a new analysis of the Sandage-Hardy sample⁵ of 60 first-rank cluster galaxies in a shell at $40-80 h^{-1}$ Mpc and claimed that no Local Group motion in the CMB dipole direction is detectable, implying, once more, a large mean velocity for the galaxies in the shell. D. Lynden-Bell (Cambridge University) presented detailed solutions for the sample of ellipticals originally reported at Hawaii. He stressed that although the solutions are fairly insensitive to whether all galaxies or only those in groups are included, they vary significantly if the sample is split into 2 hemispheres and, moreover, exclusion of a few groups in the Centaurus region causes the inferred streaming velocity to drop by 1.5 standard deviations. Thus streaming motions have been detected, but the amplitudes and directions inferred from different samples (or even from different subsamples of individual datasets) do not always agree with one another.

An alternative approach is to look directly for inhomogeneities in the galaxy distribution on the sky. O. Lahav (Cambridge University) and M. Rowan-Robinson (Queen Mary College, London) argued that optical and IRAS maps, respectively, of the galaxy distribution exhibit dipolar patterns apparently aligned with the CMB. This suggests that the source of the 600 km s^{-1} velocity of the Local Group has been localized within these maps. None of the velocity solutions, however, has an apex pointing in this direction.

What can be made of these seemingly

discordant measurements? D. Burstein (University of Arizona) argues that some of the discrepancies are caused by the different spatial distributions of the various samples. Burstein claims that, provided the same regions of space are considered, all studies agree for samples within $20 h^{-1}$ Mpc, although at greater distances significant discrepancies remain. An overall bulk-motion solution for each sample may be an oversimplification of the nearby



Computer plot by O. Lahav of the galaxy distribution in the hemisphere of sky centred on the direction of motion of the elliptical galaxy sample. This is an equal area projection based on the ESO, UGC and MCG catalogues. The galaxy clusters in Virgo (V), Centaurus (C), Hydra (H) and Antlia (A) are shown and the arrows give 90° intervals in the radial coordinate. (Courtesy of D. Lynden-Bell.)

velocity field. The galaxy distribution locally is dominated by a few superclusters that may individually have sizeable random motions. Staveley-Smith argues that the erratic behaviour of his solutions for samples of different depths can be understood in this way. S. Faber (Lick Observatory) and Burstein showed that the nearby galaxies can be neatly divided into a few patches streaming in different directions (in the regions of Ursa Major, Perseus-Pisces, Pavo-Indus and Hydra-Centaurus) or not streaming at all (Virgo southern extension). These irregular patches are typically $20 h^{-1}$ Mpc although they can be very elongated (Pavo-Indus with a long axis of $70 h^{-1}$ Mpc) and typical velocities are $400-500 \text{ km s}^{-1}$, although they can be as large as $1,000 \text{ km s}^{-1}$ (Centaurus).

When viewed in this way, it is not surprising that solutions from different samples can yield different results: the relatively large peculiar velocities of a few aggregates can conspire to give an almost arbitrary velocity for a sample which encompasses a large region of space. Each bulk-motion solution will depend on the exact sky coverage and redshift depth.

This may also explain why the apex of the motion derived from any particular sample does not agree with the optical and IRAS dipoles.

With such complex flow patterns, comparison with theoretical expectations is far from straightforward. Ideally, such comparison should take into account the detailed distribution on the sky and in redshift of each sample. N. Kaiser (Cambridge University) outlined one way in which this can be done. He used the elliptical galaxy data of Burstein *et al.*⁶ and the cluster data of Aaronson *et al.*³ to test a family of models in which peculiar velocities are generated purely by the gravitational effects of gaussian density fluctuations. This family, of which the hot and cold dark matter models are particular examples, is characterized by a coherence length and an associated velocity on that scale. In conformity with the new interpretation, Kaiser concludes that these data do not require either a large coherence length or peculiar velocities much in excess of a few hundred kilometres per second. The popular cold dark model with biased galaxy formation, previously feared to have been a casualty of large streaming motions, does not seem to be in serious conflict with the data.

At a meeting of the Royal Astronomical Society (London, 13 February 1987) Lynden-Bell presented new results from further examination of the elliptical galaxy data carried out with Faber. He singled out the large velocity of the sector of the sky within 60° of Centaurus as the main culprit in producing an apparently large bulk motion for this sample. He argued that a mass concentration of about 20 times the mass of the Virgo cluster located near the galactic plane beyond Centaurus at a distance of $46 h^{-1}$ Mpc could explain the data without requiring a net bulk motion for the sample as a whole. What is this mass concentration? Lynden-Bell says that an enhancement of galaxies near the required location can be seen in suitably oriented maps of the galaxy distribution such as the one shown in the figure. $\square$

1. Rubin, V.C., Ford, W.K. & Rubin, J.S. *Astrophys. J.* **183**, L111-L115 (1973).
2. Hart, L. & Davies, R.D. *Nature* **297**, 191-196 (1982).
3. Aaronson, M. *et al. Astrophys. J.* **302**, 536-563 (1986).
4. Bond, J.R. & van den Bergh, S. *Nature* **320**, 489-490 (1986).
5. Collins, C.A., Joseph, R.D. & Robertson, N.A. *Nature* **320**, 506-508 (1986).
6. Burstein, D. *et al.* in *Galaxy Distances and Deviations from Universal Expansion* (eds Madore, B. & Tully, R.B.) 123-130 (Reidel, Dordrecht, 1986).
7. Vittorio, N., Jusziewicz, R. & Davies, M. *Nature* **323**, 132-133 (1986).
8. Sandage, A. & Hardy, E. *Astrophys. J.* **183**, 743-757 (1973).

Carlos Frenk and John Lucey are in the Department of Physics, University of Durham, Durham DH1 3LE, UK.

Developmental genetics

Mechanisms of differentiation

James B. Hicks

THE goal of any study of cell differentiation and development is to find the primary event that causes the progeny of a single cell division to be different from one another — that is, to have different developmental potential. On page 466 of this issue¹, Amar Klar reports indirect, yet compelling, evidence that the primary signal for differentiation between daughter cells of the fission yeast *Schizosaccharomyces pombe* resides in the different structures of complementary DNA strands.

A primary developmental event must rely on a process that is independent of the regulatory cascade of gene expression that will eventually determine the differentiated phenotype. The nature of primary events is particularly critical in the early stages of development, where only one or a few cell divisions must determine multiple binary choices from a single pool of genes and cytoplasm. It is possible to imagine various mechanisms for setting different developmental programmes. In amphibian embryos, the animal and vegetal poles are apparently determined by gravity: the components of the egg cytoplasm are stratified by density and early cleavage defines cells with very different composition, thus providing a mechanism for initiating different developmental programmes in each. The structure of the DNA molecule provides another means for distinguishing progeny cells because each daughter cell receives a different complementary strand of parental DNA in each chromosome.

The structure of the DNA molecule provides another means for distinguishing progeny cells because each daughter cell receives a different complementary strand of parental DNA on each chromosome. The phenomenon of mating-type switching in sexual yeasts provides several clear examples of programmed cellular differentiation that can be studied in the absence of cell-cell interactions. These examples thus provide models for the genetically programmed portions of a developmental pathway.

The developmental event followed by Klar is the transposition of DNA segments, called cassettes, that confer the 'plus' and 'minus' sexual cell types on haploid cells. During vegetative growth, these cassettes are switched in a programmed fashion, eventually yielding progeny cells of both mating types. Pedigree studies on haploid cells show that if mating-type switching is accomplished in one half of the lineage, it will not occur in the other half. Thus, a difference in

developmental potential segregates at the first cell division of any pedigree. In diploids, however, these same lineage restrictions can be observed, but they appear to affect each of two mating-type chromosomes independently. On this evidence, Egel predicted² that the determination event is chromosomal rather than cellular.

Klar's test¹ of this hypothesis is simple and elegant. He uses the appearance of a double-stranded break in the DNA at the mating-type (*mat1*) locus that had been shown to be an early event in the switching

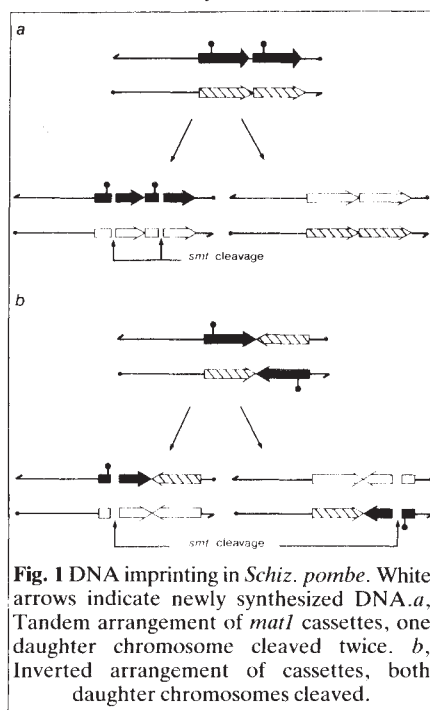


Fig. 1 DNA imprinting in *Schiz. pombe*. White arrows indicate newly synthesized DNA. *a*, Tandem arrangement of *mat1* cassettes, one daughter chromosome cleaved twice. *b*, Inverted arrangement of cassettes, both daughter chromosomes cleaved.

process as a physical marker for the developmental capacity to switch. The site of this break (designated *smt*) is unique and is located within the segment of *mat1* DNA that becomes rearranged during switching. By placing two *mat1* cassettes on the same chromosome in tandem or inverted orientation, Klar could compare the number of cells containing breaks at one or both sites. The clear result is that when the cassettes are installed in tandem, any cell containing a double-stranded break is cleaved at both *smt* sites (Fig. 1*a*).

In the inverted orientation, the individual *smt* sites are cleaved at the same frequency, but both sites are never cleaved in the same cell (Fig. 1*b*). The conclusion drawn by Klar is that at each division one of the two complementary strands comprising the *mat1* locus is asymmetrically marked (perhaps by hemi-methylation) and that this mark is a signal that determines the response of the daughter chro-

mosome to the *smt* endonuclease and hence its ability to switch. He has termed this process DNA imprinting. Although the details are still unknown, Klar's work is the first demonstration that such a mechanism plays a role in cellular differentiation.

A very different story is unfolding in the case of mating-type switching in the budding yeast *Saccharomyces cerevisiae*. In this organism, not only can pedigrees be followed but mother cells can be distinguished from their daughters during the budding process.

Studies of individual dividing and switching cells have yielded rules for this simple developmental process^{3,4}. One of these is that in any cell division only the mother cell has the potential to switch mating type — a clear segregation of developmental potential. The components of the switching system in budding yeast are well known. Switching occurs by transposition of DNA cassettes from silent to expressed chromosomal locations and transposition is catalysed by an endonuclease, the product of the *HO* gene. But why should switching occur only in mother cells if the same genetic components are present in both mothers and buds?

In 1983, Nasmyth showed⁵ that the distinction between mother and bud results not from DNA marking, but from the cell-cycle-specific transcriptions of the *HO* gene and the localization of the *HO* gene product to mother cells. This was clearly an important discovery, but only set the question back a step. What causes the *HO* product to be localized? The answer may lie in an interlocking circuit of regulatory gene expression paced by the cell cycle.

Classical mutant-suppressor studies have uncovered two gene products that act as positive and negative regulators of *HO* transcription. Particularly noteworthy is a repressor of *HO* described in two papers published last month in *Cell* by a group led by Ira Herskowitz⁶ and by Nasmyth *et al.*⁷. This repressor, called *SIN3* by Herskowitz's group and *SDII* by Nasmyth's group, is apparently responsible for preventing *HO* expression in daughter cells. Furthermore, both groups report that another *HO* regulator, *SWI5*, is an antagonist of *SIN3*. So *HO* is expressed in the presence of both products.

Herskowitz's group⁶ proposes a model for the asymmetrical regulation of *HO* that does not rely on another factor for localization of *SIN3* or *SWI5* and thus is independent of the pathway. The model makes use of the fact that a separate regulatory circuit allows *HO* to be turned on only during a small portion of the cell cycle just before DNA synthesis^{7,8}. Furthermore, daughter cells normally spend more time in the G₁ phase of the cell cycle than mother cells, presumably responding to yet another signal requiring a certain cell size to progress through the cycle. Under

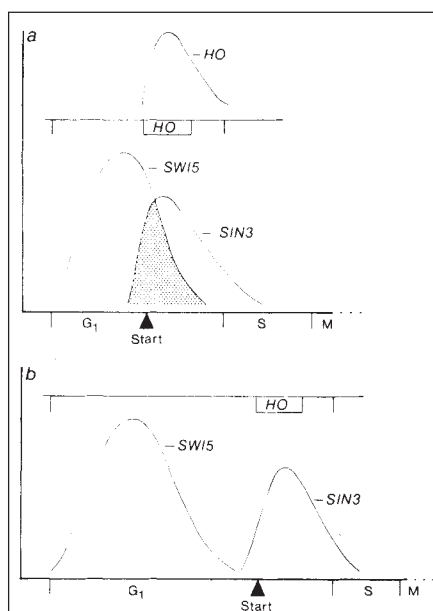


Fig. 2 Cell-cycle-dependent model for mother-daughter differentiation of *HO* gene in *Sacc. cerevisiae*. *a*, In mother cells, *SWI5* expression overlaps *SIN3* expression during G_1 . Shaded area, inactivation of *SIN3* during 'start', thus allowing expression of *HO*. *b*, In daughter cells, *SWI5* decays before start, *SIN3* represses *HO* and no switching occurs.

the terms of the Sternberg model, the *SWI5* and *SIN3* products are also cell-cycle regulated, *SWI5* being expressed early in G_1 and *SIN3* later, near the start of the S phase of the cycle (Fig. 2a). In mother cells, progressing swiftly through G_1 , both molecules would be present, *SIN3* would be inactivated and *HO* would be transcribed. In daughter cells, however, sufficient time would be spent in mid- G_1 phase for the *SWI5* product to decay before the window for *HO* and *SIN3* expression opens. In the absence of *SWI5*, *HO* is repressed and daughter cells do not switch (Fig. 2b).

The gratifying property of this model is that, given a cellular process independent of the developmental pathway under observation (in this case the timing of cell division itself) it provides a mechanism for a primary event in the differentiation of cell types rather than simply setting the question back another step. Testing the model will depend on cloning the *SWI5* and *SIN3* genes and testing their timing during the cell cycle. We probably will not have long to wait for the answer. □

1. Klar, A. *Nature* **326**, 466–470 (1987).
2. Egel, R. *Curr. Genet.* **8**, 205–210 (1984).
3. Hicks, J. B. & Herskowitz, I. *Genetics* **83**, 245–258 (1976).
4. Strathern, J. N. & Herskowitz, I. *Cell* **17**, 371–381 (1979).
5. Nasmyth, K. A. *Nature* **303**, 670–676 (1983).
6. Sternberg, P. W., Stern, N. J., Clark, I. & Herskowitz, I. *Cell* **48**, 567–577 (1987).
7. Nasmyth, K., Stillman, D. & Kipling, D. *Cell* **48**, 577–587 (1987).
8. Breeden, L. & Nasmyth, K. *Cell* **48**, 389–397 (1987).

James B. Hicks is in the Department of Molecular Biology at the Scripps Clinic and Research Foundation, La Jolla, California 92037-1093, USA.

Meteorites

Nearer yet farther away

Colin Pillinger

CONTROVERSY about the origin of the organic matter in meteorites has existed almost since it was first accepted that meteorites are stones fallen from the sky. A new study by Epstein *et al.* on page 477 of this issue¹ suggests that certain groups of compounds have been synthesized outside the Solar System, so extractable species join macromolecular material and elemental dust as presolar phases.

The Alais meteorite was one of the first dozen or so well-documented falls that helped persuade scientists, at the turn of the eighteenth century, that extraterrestrial material was indeed reaching Earth. This sample, analysed² by Berzelius in 1834 using destructive distillation, yielded a blackish substance, water, carbon dioxide and ammonia, and gave the first clear hint that amino acids would eventually be found regularly in carbonaceous chondrites. Berzelius recognized that the organic matter from Alais was different from terrestrial humus. Possibly because he wanted to be in vogue with contemporary hypotheses that meteorites came from the Moon, which he knew lacked sufficient atmosphere to support life, he stated that there was nothing in his observations to warrant a conclusion that "organic nature" existed on other celestial bodies.

Others who followed Berzelius were less conservative³. Wohler, Harris and Cloez, who were responsible for the analyses of the Kaba, Cold Bokkeveld and Orgueil stones, all believed that some vital force had been involved in production of their organic matter — Helmholtz and Thompson (later Lord Kelvin) were also convinced. Only Berthelot, who had successfully synthesized hydrocarbons in

the laboratory in 1858, believed that a similarity of structure did not constitute an argument for a biological origin.

Although there was a great desire to know more about the nature of meteorite organics, the lack of techniques appropriate to the problem prevented any progress for almost 100 years. By that time gas chromatography and mass spectrometry had found their way into the hands of geochemists interested in the structure of sedimentary organic matter and the origin of petroleum. In 1961, to effect a resurrection of the problem, Nagy, Hennessey and Meinschein reported⁴ mass-spectral data for carbonaceous species from Orgueil. Their identifications suggested that biogenic processes existed beyond the Earth. The new methodology very rapidly revealed⁵ that carbonaceous chondrites contain alkanes, aromatics, carboxylic acids, amino acids and a host of other compound classes — in fact virtually everything that might be encountered in terrestrial sediments.

Many were sceptical of these results, believing the analyses were detecting terrestrial contamination. After all, some of the meteorites being studied, such as Orgueil (fell 1864) and Cold Bokkeveld (fell 1838), had lain on museum shelves and been lovingly handled by countless admirers. But no less a personage than Harold Urey wrote⁶ "If the materials were of terrestrial origin, it would be firmly suggested that they were of biological origin and indigenous to the samples". Rather surprisingly, Urey, the father of stable isotope geochemistry, did not cite any isotope data to support his argument that the organic matter was indigenous, although the evidence had been available

Five stamps, designed by David Hartley, were issued on 30 March by Ascension Island to mark the 25th anniversary of the first United States Earth orbit. On this flight, launched on 20 February 1962, John Glenn orbited Earth three times in the capsule Friendship 7. The two stamps not shown depict lift-off (18p) and re-entry (25p). (Courtesy of Crown Agents Stamp Company Ltd, Surrey SM1 4RN, UK.)

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Donald MacKay (1922–1987)

DONALD MacKay, who died on 6 February after a long illness, was one of the pioneers of theoretical neurobiology. A physics graduate of St Andrews University, with 3 years' wartime experience in radar research, he joined the physics department at King's College London in 1946, where he turned his attention to high-speed analogue computers and their possible relevance to the human brain. He was one of the first British scientists to appreciate the importance of Shannon's information theory and Wiener's cybernetics in setting theoretical limits to the performance of the central nervous system, and to see the importance of neural models such as that of McCulloch and Pitts, whose work he much admired.

The decisive event in his career was his move to Keele University in 1960, to found the Department of Communication and Neuroscience, now a research institute of international standing. He assembled there a team of psychophysicists, neurophysiologists and electronic engineers, and together they attacked some of the more intractable problems of vision and hearing, such as the origin of visual after-effects, the mechanisms of speech perception and the processes responsible for visually evoked brain potentials.

But Donald MacKay is probably best known for his contributions to the philosophical debate about free will and determinism, and the unitarity of human consciousness. In many books and public lectures he argued that the predictability of a person's choices (accepting the determinist hypothesis for the sake of argument) does not constitute a violation of that person's freedom to choose, because telling the person about one of his choices in advance would alter his cognitive state in a way that could not, in principle, be allowed for in making the prediction: one would have to

allow for the effect of allowing for the effect . . . of doing so. And once it is admitted that a prediction could be undermined by revealing it, its objective validity can immediately be questioned. This argument, and later variants designed to meet specific objections, established MacKay as a first-rate philosopher of science.

A further important contribution to the philosophy of neuroscience was MacKay's recent discussion of the behaviour and reported experiences of split-brain patients. Here the question was whether the available evidence needed to be interpreted as demonstrating the existence of two independent minds controlling the same body. MacKay's analysis is both sensitive and acute, and arrives at a negative conclusion.

Apart from his own research, MacKay was an effective communicator of scientific and philosophical ideas. He was a frequent attendee at meetings of the Neurosciences Research Program at the Massachusetts Institute of Technology, and one of its wisest counsellors. His gentle insistence on the categorical distinction between 'I-language' and 'it-language' in discussions of the mind and the brain was often, however, too subtle for his hard-nosed colleagues, who felt more at home with 'wet' neuroscience (as they called it) than with the 'dry' variety that MacKay seemed to be peddling.

Donald MacKay was an active churchman and a vigorous Christian apologist, but he never allowed his religious convictions to short-circuit his scientific thinking. His high standards of critical rigour will be sadly missed in the debate on science and religion.

H.C. Longuet-Higgins

H.C. Longuet-Higgins is Royal Society Research Professor in the Laboratory of Experimental Psychology, University of Sussex, Brighton BN1 9QG, UK.

for some time.

In many ways, the year 1969 was a watershed for meteoritics: for example, the fall of the Murchison type CM2 carbonaceous chondrite provided an abundant, pristine sample for organic analysis. By 1970, it was possible to separate amino-acid enantiomers using capillary gas chromatography. Thus, Kvenvolden *et al.*⁸, investigating extracts of Murchison, could show that its amino acids were racemic mixtures of the R and S forms. Moreover, some of the compounds identified were not among the 20 or so amino acids that commonly constitute biogenic proteins. Their conclusion was simple: they had the first irrefutable evidence that the amino acids must be extraterrestrial and possibly abiogenic. A prodigious effort by several investigators (see ref. 9 for a review) has increased the number of amino-acid structures identified in

Murchison to more than 50. The diversity of this collection, together with the knowledge that the simplest molecules (glycine, aminobutyric acid and so on) are the most abundant, argue strongly in favour of a synthesis from small-molecule precursors.

Until relatively recently two types of synthetic route, Miller–Urey reactions¹⁰ or Fischer–Tropsch processes¹¹, had been proposed for converting gaseous mixtures variously composed of H₂, H₂O, CH₄, CO and NH₃ into far more complicated molecules. Tacitly it was assumed that the events involved took place in the early Solar System, possibly on the meteorite parent bodies. The discovery¹² of increasingly complex species in dark interstellar clouds, however, suggests an alternative, especially as links have been made between water released by step heating of bulk primitive meteorites or concentrated acid residues^{13–16} and the interstellar

clouds. Ion-molecule reactions^{13,17} at low temperature in the clouds are seen as possibly the only feasible explanation for the 'astronomical' deuterium enrichments (up to a factor of six)¹⁸ measured for the water.

In their paper in this issue¹, Epstein and Krishnamurthy from Caltech, in collaboration with Cronin and colleagues from Arizona State University, take this hypothesis a step further. They suggest that amino and carboxylic acids are among the carriers of this characteristic deuterium signature and of an unusual nitrogen isotope composition, which thus pinpoint interstellar clouds outside the Solar System as their provenance. Their theory is based on isotope analyses conducted at Caltech on extracts from the Murchison meteorite that the Arizona State group (vastly experienced in meteorite amino-acid studies) would consider to be appropriate for amino and monocarboxylic acids.

As with any finding of this type, there has to be a certain amount of caution. First, the yield of hydrogen from the meteorite compounds is less than 100 per cent because it was not realized until too late that the recovery procedure did not completely transfer the water produced when the samples were burned before isotope measurement. The authors argue that because of this (together with some exchange reactions incompletely corrected for and possible contamination by the ion-exchange column material used in separating the compounds) their data provide only minimum deuterium abundances. They do not discuss the possible counterbalance of fractionation effects, presumably because they were not seen in procedural blanks.

Second, and to me more worrying, the analysis of the elements for the total amino-acid fraction is C₇H₈N. Because it is known that by far the most abundant amino acids are glycine (C₂H₅NO₂), aminobutyric acid (C₄H₉NO₂), alanine (C₃H₇NO₂), glutamic acid (C₅H₉NO₂), proline (C₅H₉NO₂) and valine (C₆H₁₁NO₂), the fraction analysed is clearly nitrogen deficient. Experimental checks on isoleucine revealed no problems with technique which therefore implies the presence of a carbon-, hydrogen- and possibly oxygen-containing species, either indigenous to the meteorite or a contaminant. I would almost prefer the latter because this would not change the interpretation. But the investigation of isotope anomalies in meteorites has frequently been likened to studying a nest of Russian dolls — each layer is stripped away to reveal another. Based on this analogy, the high deuterium value could be associated with a tiny amount of something else which coincidentally co-exists in the amino-acid fractions.

I find the production of amino acids in interstellar clouds highly plausible but, like the authors, I look forward to the isotope analysis of a single separated com-

pound to provide stronger proof. This sort of effort is essential to provide 'ground truth' to the astronomical observations and to pave the way for the analysis of cometary samples — if we eventually obtain them. □

1. Epstein, S. *et al.* *Nature* **326**, 477–479 (1987).
2. Berzelius, J.J. *Ann. Phys.* **33**, 113–148 (1834).
3. Burke, J.G. in *Cosmic Debris* (University of California Press, 1986).
4. Nagy, B., Hennessy, D.J. & Meinschein, W.G. *Ann. N.Y. Acad. Sci.* **93**, 25–35 (1961).
5. Hayes, J.M. *Geochim. cosmochim. Acta* **31**, 1395–1440 (1967).
6. Urey, H.C. *Science* **151**, 157–166 (1966).
7. Briggs, M.H. *Nature* **197**, 1290 (1963).
8. Kvenvolden, K. *et al.* *Nature* **228**, 923–926 (1970).

9. Cronin, J.R. & Pizzarello, S. *Geochim. cosmochim. Acta* **50**, 2419–2427 (1986).
10. Miller, S.L. & Urey, H.C. *Science* **130**, 245–251 (1959).
11. Hayatsu, R. & Anders, E. *Topics Current Chem.* **99**, 1–39 (1981).
12. Winniewisser, G. *Topics Current Chem.* **99**, 39–72 (1981).
13. Kolodny, Y., Kerridge, J.F. & Kaplan, I.R. *Earth planet. Sci. Lett.* **46**, 149–158 (1980).
14. Robert, F., Merlivat, L. & Javoy, M. *Nature* **282**, 785–789 (1981).
15. McNaughton, N.J. *et al.* *Nature* **294**, 639–641 (1981).
16. Robert, F. & Epstein, S. *Geochim. cosmochim. Acta* **46**, 81–95 (1982).
17. Geiss, J. & Reeves, H. *Astr. Astrophys.* **18**, 126–132 (1981).
18. McNaughton, N.J., Fallick, A.E. & Pillinger, C.T. *J. geophys. Res.* **87**, A297–A302 (1982).

Colin Pillinger is in the Department of Earth Sciences, Open University, Milton Keynes MK7 6AA, UK.

Archaeology

A Neolithic way of death

Patricia Phillips

ARCHAEOLOGISTS recover either deliberately deposited material or litter. Deliberate deposition is relatively rare: events such as formal burials or the hoarding of weapons or jewellery come immediately to mind, and Palaeolithic cave paintings and Bronze Age rock engravings were also deliberately placed. Another exception to this rule is described in recent reports by P. Villa and collaborators^{1–3} showing that early Neolithic visitors to a south-east French cave butchered both animal and human bodies (Fig. 1), an interpretation made possible because the joints and butchering tools were deliberately discarded in pits and were not subsequently disturbed.

Surface surveys have become more popular than excavations in the past decade, both for economic and for theoretical reasons. Surface survey is cheaper than excavation; it can be effective, either alone or with aerial photography, for studying landscape usage at different periods and in different areas; and it is non-destructive. But any archaeological material found on the surface has usually been brought up by ploughs and is no longer in place. Even if it formed a closed find until recently, once ploughed up it is disturbed and, out of context, is litter. Partly as a

result of the increase in surface survey, as well as the need to interpret surface finds, much effort has been expended in defining the parameters for interpreting ancient litter (for example, ref. 4).

In contrast, the south-east French cave of Fontbrégoua, Salernes, Var, offers clear examples of deliberate gestures from the past^{1,2}. This cave has extensive deposits dating at least from the Upper Palaeolithic, but the activities I refer to here date from the earlier phases of the Neolithic period, in the fifth and fourth millennia BC. Precise excavation and careful recording allow a detailed picture of several events. Because groups of bones lie in anatomical connection, the chemical wear on the bones is identical and the occupation level surrounding the pit contains hardly any bone, it is clear that the deposits have not been disturbed. The pit contents vary, one (feature 2) comprising a circle of

stones around part of a domestic ox skull — similar, possibly ritual, deposits occur in other contemporary cave sites. The most spectacular deposit (feature H3) is a pit containing human bones from six individuals (three adults and two children, one of uncertain age), together with two stone bracelets in fragmentary condition, and a small axe (Fig. 2). The skulls, some leg bones and shoulder blades are missing; the axe is believed to have been used in butchering the bodies. There were no skulls in the pit, but the skulls have been found in a more disturbed part of the site (H1). The other eight pits contain the remains of butchered animals, including a nearly complete sheep carcass. The lower leg bones and shoulder blades were missing, but otherwise this act of butchering must have taken place in the cave. In another case, incomplete skeletons of three wild boar were found with flint blades and a small axe. Microwear analysis showed up the polish on the edge of the blade from cutting meat. It is likely that the boar were transported whole to the cave for butchery and consumption. The similarity of butchering techniques on human and animal bones is evidence for

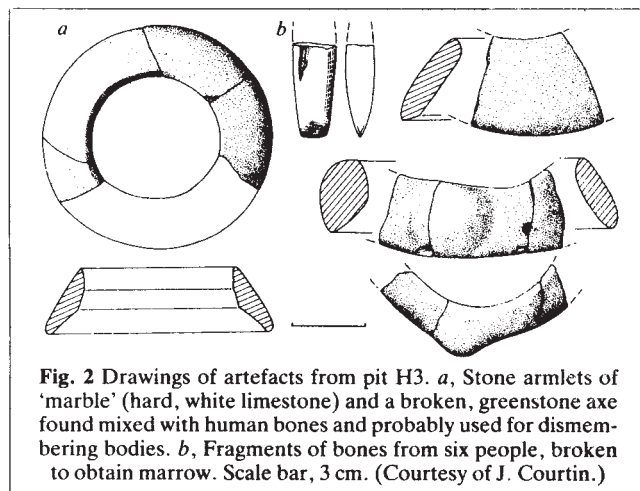


Fig. 2 Drawings of artefacts from pit H3. a, Stone armlets of 'marble' (hard, white limestone) and a broken, greenstone axe found mixed with human bones and probably used for dismembering bodies. b, Fragments of bones from six people, broken to obtain marrow. Scale bar, 3 cm. (Courtesy of J. Courtin.)

early Neolithic cannibalism.

Perhaps archaeologists should again concentrate on deliberate deposits. Whether on a small scale, as in the examples I have described, or on a large scale, such as a Viking ship burial, these deposits are the outcome of intentional behaviour. As the Fontbrégoua cave shows, excavation and analysis of deliberate deposits can re-create past activities with precision. Although surface surveys are ideal for picking up large-scale patterns of previous land use, our knowledge of prehistory depends ultimately on the information provided by the deliberate deposit. □

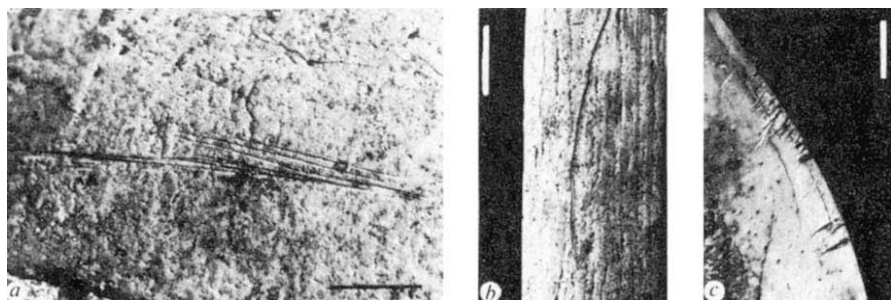


Fig. 1 Fontbrégoua's cave. a, Skinning cut marks from a flint knife on the frontal bone of a child's skull, from pit H3. b, Filleting cut marks on a human tibia, from pit H1. c, Dismembering cut marks on the edge of a child's clavicle, from pit H1. Scale bars, 1 cm. (Courtesy of J. Courtin and P. Villa.)

1. Villa, P., Helmer, D. & Courtin, J. *Bull. Soc. Préhist. Française* **82**, 389–421 (1985).
2. Villa, P. *et al.* *Science* **233**, 431–437 (1986).
3. Villa, P. *et al.* *Gallia Préhist.* **29**, 143–171 (1986).
4. Schiffer, M. *Am. Antiq.* **48**(4), 675–706 (1983).

Patricia Phillips is in the Department of Archaeology and Prehistory, University of Sheffield, Sheffield S10 2TN, UK.

Cytoskeleton

Role of fodrin in secretion

Robert D. Burgoyne and Timothy R. Cheek

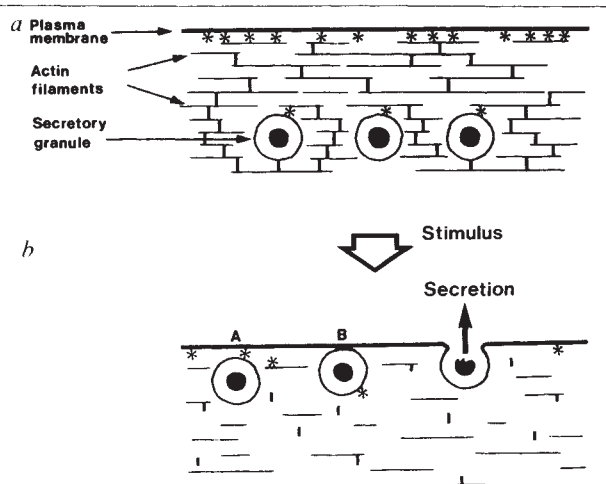
A REPORT by Perrin, Langley and Aunis on page 498 of this issue¹ suggests that the cytoskeletal protein fodrin is involved in secretion. In many types of cells, secretion requires an external signal which, through a second messenger, leads to movement of secretory granules or vesicles to the plasma membrane, fusion of the granule and membrane and consequent release of granule contents by exocytosis. Like all other organelles, secretory granules are embedded in a cytoskeletal meshwork and to reach the plasma membrane they have to pass through the network of actin filaments (F-actin) that lies beneath the plasma membrane.

It has often been speculated that secretory granules are actively transported by an actin-myosin-based contractile system, by analogy to the sliding filament mechanism of muscle contraction. But the most likely significance of the cytoskeleton, particularly the actin network, is in preventing exocytosis²⁻⁸ until the appropriate signal is received by the cell. In support of this idea it has been shown that agents causing disassembly of the F-actin network, such as cytochalasins, can promote calcium-stimulated secretion from permeabilized adrenal chromaffin cells³ and can lower the calcium dependency of secretion from neutrophils⁴, whereas phalloidin, an actin-filament stabilizer, inhibits secretion⁵. The disruption of the actin network (see figure) thought to occur in response to external signals could involve regulation of calcium-dependent actin-crosslinking proteins such as caldesmon^{5,6} and/or depolymerization of F-actin, which occurs following stimulation of adrenal chromaffin cells⁸.

The new work reported in this issue¹ demonstrates that an antiserum against the α -subunit of brain fodrin, but not antisera against filamin and vinculin (other cytoskeletal proteins), partially inhibits secretion elicited by calcium from digitonin-permeabilized adrenal chromaffin cells. Fodrin and its erythrocyte coun-

terpart spectrin are found in the cytoplasm of cells and are thought to be involved in the linkage of F-actin to the plasma membrane. In the adrenal chromaffin cell fodrin is mainly present in the cell cortex¹ but has also been detected on the isolated membranes of the secretory granules⁹.

What role for fodrin in secretion do the new results suggest? The fact that the antiserum against α -fodrin inhibits secretion could mean that fodrin is important in mediating interactions between F-actin



Organization of the cytoskeleton and proposed changes leading to secretion. *a*, In resting cells, secretory granules are cross-linked within the cytoskeleton and are prevented from reaching the plasma membrane by the dense crosslinked F-actin network in the cell cortex. The stars represent actin filaments attached to the plasma membrane and secretory granules through fodrin. *b*, Following stimulation the secretory granules are released from the cytoskeleton and the F-actin network is disassembled. Some clearing of fodrin from the plasma membrane occurs but interaction between the secretory granules and plasma membrane could be mediated by fodrin remaining on the plasma membrane (A). Alternatively, secretory granules might interact with the plasma membrane only at sites where fodrin has been removed (B). Vertical lines represent caldesmon and other actin-crosslinking proteins.

and secretory granules at increased calcium concentrations. This possibility is unlikely because it has previously been demonstrated¹⁰ that micromolar concentrations of calcium inhibit F-actin binding to secretory granule membranes. Another possibility is that fodrin is directly responsible for the initial interaction of secretory granule and plasma membranes leading to membrane fusion (see figure); further work is needed to test this speculation. Finally, a clue for another explanation of the inhibitory effect of the antiserum comes from the earlier finding¹¹ of Perrin and Aunis that stimulation of adrenal chromaffin cells results in loss of fodrin from parts of the cell. Fodrin is highly susceptible to calcium-activated protease

activity, but this can be prevented in rat brain fodrin by prior incubation of brain membranes with an anti-fodrin antiserum¹². Thus, calcium-activated proteolysis of fodrin in chromaffin cells may be part of the process by which the peripheral cytoskeleton is cleared to allow secretory granules access to exocytotic sites, and it is this mechanism that the antiserum inhibits. Further work should determine the exact way in which the antiserum inhibits secretion.

One aspect of the results with α -fodrin that remains to be explained is that the antiserum produces only a 50 per cent inhibition of secretion. Perrin *et al.* propose¹ a plausible explanation that the uninhibitable stage of secretion is caused by exocytosis of those secretory granules already at the plasma membrane.

Another intriguing aspect of the control of the cytoskeleton in secretory cells is the possible importance of a guanine-nucleotide-binding (G) protein. It has been suggested¹³ that GTP analogues stimulate secretion from neutrophils by acting on a G protein that directly controls exocytosis. The GTP analogues may be acting on the cytoskeleton: the myosin and actin associated with isolated neutrophil membranes, presumably as part of the cell cytoskeleton, are solubilized from the membranes by incubation with GTP- γ -S (ref. 14). The exact relationship of this effect of GTP- γ -S to the events occurring during secretion remains to be determined.

The nature of the events involved in the membrane fusion leading to exocytosis is still a complete mystery. From early ideas that calcium directly mediates membrane fusion, it has now become apparent that second messengers may regulate several different steps leading to exocytosis. The experimental approach described by Perrin *et al.* may be useful in elucidating the mechanism of exocytosis. □

1. Perrin, D., Langley, O.K. & Aunis, D. *Nature* **326**, 498–501 (1987).
2. Orci, L., Gabbay, K.H. & Malaisse, W.J. *Science* **175**, 1128–1130 (1972).
3. Lelkes, P.I., Friedman, J.E., Rosenheck, K. & Oplatka, A. *FEBS Lett.* **208**, 357–363 (1986).
4. Lew, D.P. *et al.* *J. Cell Biol.* **102**, 2197–2204 (1986).
5. Burgoyne, R.D., Cheek, T.R. & Norman, K.M. *Nature* **319**, 68–70 (1986).
6. Sobue, K., Tanaka, T., Kanda, K., Ashino, N. & Kakiuchi, S. *Proc. natn. Acad. Sci. U.S.A.* **82**, 5025–5029 (1985).
7. Bennett, J. & Weeds, A. *Br. med. Bull.* **42**, 385–390 (1986).
8. Cheek, T.R. & Burgoyne, R.D. *FEBS Lett.* **207**, 110–114 (1986).
9. Aunis, D. & Perrin, D. *J. Neurochem.* **42**, 1558–1569 (1984).
10. Fowler, V.M. & Pollard, H.B. *Nature* **295**, 336–339 (1982).
11. Perrin, D. & Aunis, D. *Nature* **315**, 589–592 (1985).
12. Siman, R., Baudry, M. & Lynch, G. *Nature* **313**, 225–228 (1985).
13. Barrowman, M.M., Cockcroft, S. & Gomperts, B.D. *Nature* **319**, 504–507 (1986).
14. Huang, C.-K. & Devanney, J.F. *FEBS Lett.* **202**, 41–44 (1986).

Robert D. Burgoyne and Timothy R. Cheek are in the MRC Secretory Control Group, the Physiological Laboratory, University of Liverpool, PO Box 147, Liverpool L69 3BX, UK.

Defending Mendel merely perpetuating a myth

SIR—Though I agree with Van Valen¹ that “Mendel was no fraud”, my own recent enquiry, “Are Mendel’s results really too close?”², led me to conclude that they are. After reviewing all the evidence and reanalysing the entire data I found “that Mendel’s data exhibit two independent peculiarities both of which point to the same conclusion, namely, that in general the segregations agree more closely with what Mendel expected than chance would dictate. Where Mendel expected a 2:1 ratio even though the true expectation must have been different it was the former ratio which was achieved, and throughout the rest of his results there is a persistent lack of extreme deviations”.

No serious commentator with a first-hand knowledge of Mendel’s paper has ever suggested that Mendel actually falsified his data. Yet this popular charge is being strengthened by the frequent appearance in the scientific literature of papers of doubtful value intended by their well-meaning authors to deflect it.

Van Valen’s letter¹ is no exception. He quotes four papers in support of his view that Mendel did not falsify his data. Contrary to Van Valen’s assertion, the first of them (here given its correct reference³) contains no analysis of Mendel’s data, but rather a brief well-referenced review of the problem concluding with the extremely cautious opinion “one tends towards the impression that Fisher’s original conclusions were somewhat extreme: there was probably no deliberate attempt at mendacity on Mendel’s part”. (Fisher, of course, had not suggested that there had been).

The second⁴ and third⁵ papers cited by Van Valen are entirely irrelevant because they do not treat the goodness-of-fit problem but the quite separate issue of Mendel not finding linkage, whilst the fourth⁶ contains substantial errors, as detailed in my review⁷.

Gregor Mendel does not need the sort of defence which is being mounted. His paper is one of the greatest products of the human mind even if we still do not fully understand it. “There is every reason for each student of Mendel to keep an open mind and to make his own analysis and form his own conclusion”⁸. Mine is that Mendel’s results really are too close, but certainly not that he falsified them.

A. W. F. EDWARDS

Gonville & Caius College,
Cambridge CB2 1TA,
UK

1. Van Valen, L. M. *Nature* **325**, 395 (1987).

2. Edwards, A. W. F. *Biol. Rev.* **61**, 295–312 (1986).

3. Pigorsch, W. W. *Hist. Sci.* **24**, 173–182 (1986).

4. Blixt, S. *Nature* **256**, 206 (1975).

5. Douglas, L. & Novitski, E. *Heredity* **38**, 253–257 (1977).

6. Monaghan, P. & Corcos, A. J. *Hered.* **76**, 307–309 (1985).

How many Chernobyl fatalities?

SIR—I am afraid that Robin Russell Jones’ underestimates the number of probable cancer fatalities as a result of the Chernobyl accident because he does not take into account the results of several investigations that are less well known, but by no means to be taken less seriously, than those on which he relies.

Taking into account both the dose reassessment by Loewe² and a more correct control group than the traditional 0–9 rad group, the life-time cancer risk per rem per million persons has been determined from atom bomb victims to be 300–1,800 fatalities³. Gofman derives the number 3,000–4,000 from the same data by careful consideration of age dependencies and other currently neglected factors⁴. Using data from the Hanford workers Mancuso *et al.* derived a supra-linear dose effect curve of square-root-shape, which rises steeply in the low dose range (< 1 rem) and approached a lower slope, consistent with the findings from atom bomb victims and medical exposures in the higher dose range (> 20 rem)⁵. From this, Morgan⁶ calculated 7,000 fatal cancers per rem per million persons in the low dose range.

For the 60 million inhabitants of West Germany collective dose assessments range from 9 million to 72 million person rems, implying a range of about 10,000–500,000 cancer fatalities, with a most probable value, in my judgement, of around 250,000.

But cancer is not the only fatal consequence of ionizing radiation. Lave *et al.*⁷ have reported a correlation between mortality and the amount of ¹³⁷Cs and ⁹⁰Sr in milk. From their work one can derive an overall formula that a diet characterized by 1 Bq of ¹³⁷Cs or ⁹⁰Sr per litre implies an increase of mortality of 60 or 400 per million people a year, respectively. As the method of Lave *et al.* mainly assessed short time relationships, one must suspect that the mortalities are largely not due to cancer but to infectious diseases that result from the immune suppression, following relatively soon after irradiation. Qualitative correlations between ⁹⁰Sr in the diet and mortalities from several infectious diseases can also be derived from US vital statistics⁸, and Mehning⁹ has found correlations between fall-out and a number of blood irregularities, like leukocyte depression. The reason for these effects may be reduction of immunological competence by ionizing radiation. That low level irradiation of bone marrow by ⁹⁰Sr may have a supra-linear effect on the immune system is supported by an early observation of Stokke¹⁰ that the cellularity depression in bone marrow of mice shows just such a supra-linear dependency on ⁹⁰Sr dose.

From Lave *et al.* and the measured ¹³⁷Cs data in West Germany, the country should suffer 100,000–200,000 mainly non-cancer fatalities which should be added to the number of cancer fatalities derived from the dose effect curve described above. For the UK population investigated by Fry *et al.*¹¹ one is similarly forced to add 1,000–2,000 to obtain the total number of fatalities due to Chernobyl.

JENS SCHEER

Physics Department,
University of Bremen,
Postfach, 2800 Bremen 33, FRG

1. Jones, R. R. *Nature* **323**, 585–586 (1986).

2. Loewe, *et al.* Lawrence Livermore National Laboratory D80-14.

3. Carbonell, P. & Schmitz-Feuerhake, I. IAEA SM-266/23, 45 (1983).

4. Gofman, J. National Meeting of the American Chemical Society (Anaheim, September 1986).

5. Mancuso, T., Stewart, A., Kneale, C. W. *Br. J. ind. Med.* **38**, 156 (1981).

6. Morgan, K. *Bull. atom. Sci.* **34**(4), 30 (1978).

7. Lave, L., Leimhardt, S., Kaye, M. *Report W.P. 19-70-1* (Graduate School of Industrial Administration, Carnegie Mellon Univ., Pittsburgh, July 1971).

8. *Summary of Vital Statistics* (Department of Health, Albany, New York, 1983).

9. Mehning, C. E. *Protectio vitae* **65**, 220 (1972).

10. Stokke, T. *et al. Acta radiologica* **7**, 321 (1968).

11. Fry, F. A. *et al. Nature* **321**, 193 (1986).

Cause of death of the American Indians

SIR—The American Indians suffered a demographic disaster after the discovery of the New World in 1492, which has usually been blamed on smallpox. But as this disease was not introduced into Santo Domingo until 1518, by when the population of that island had decreased from 1,100,000 (in 1492) to only 10,000, smallpox cannot be blamed. Other suggestions have never been properly documented. From an analysis of contemporary evidence from Columbus and his physician Alvarez Chanca, supported by evidence from Martyr Anglerius, Fernandez de Oviedo, Bishop Las Casas (whose father and uncles went through the epidemic), Columbus’s son Hernando, and Herrera y Tordesillas, I would like to suggest that the greatest mortality was caused by an epidemic of swine influenza.

The Santo Domingo epidemic started at La Isabela, the first city of the New World, on 9 December 1493, the day after 1,500 men and domestic animals had landed from the seventeen ships of Columbus’s second voyage. Domestic animals, including eight sows, had been taken on board the Admiral’s ship at La Gomera in the Canary Isles on 5–7 October 1493, but contact between these animals and expedition members took place only after unloading, when the horses were reported to be completely “lost” from illness.

All sources agree on the place, date, clinical manifestations and mortality of the epidemic. The clinical picture was of an acute infectious disease that was extremely contagious and immediately

affected everyone in the expedition, including the Admiral. Its symptoms were high fever, ague, prostration and great mortality, though those who recovered did not relapse. Reports of other outbreaks of the disease, which followed invasion of the mainland in 1514–19, mention respiratory involvement and epistaxis. The sources consulted indicate that after affecting the Spaniards, the Indians started to die “in infinite numbers”.

The short incubation period of the 1493 epidemic, and its evolution rule out the possibility of the illness being malaria or yellow fever, as has been suggested. But, both clinically and epidemiologically the disease can be identified as influenza.

In a study of the role of swine influenza viruses in pandemics of human influenza¹, I have compared the demographic evolution of the Antilles since the arrival of Columbus in 1492 with that of the Philippines since the arrival of Magellan in 1521. Both archipelagos have a similar geographic extension and climate. But whereas the pre-Columbian Indians of the Antilles

had practically no domestic animals and so were exposed to domestic animal viruses only when Columbus landed, the nation of the Philippines had domestic animals including at least three species of pigs before Magellan arrived.

Therefore the inhabitants of the Philippines had an acquired immunity that enabled them to withstand their immunological collision with the Spaniards while the Antilles Indians perished in very large numbers for lack of immunity. In their own way, the chroniclers were aware of the Indians' selective immunity. Fernandez de Oviedo, for instance, admired Indian resistance to yaws and venereal disease, while Bishop Las Casas noticed their fragility to respiratory ailments. Solórzano Pereira was perhaps the most clear-sighted when he stated that the breath of other people killed the Indian.

FRANCISCO GUERRA

*Department of History of Medicine,
Universidad de Alcalá de Henares,
Madrid, Spain*

1. Guerra, F. *Revista de Indias*, Madrid 176, 325–347 (1985).

Importance of satellites for weather forecasting

SIR—Your article¹ entitled “Satellite weather around the corner” fails to do justice to the vital importance of satellite observations as currently employed in operational weather forecasting and presents a too pessimistic and limited view of their future importance.

The article was based on a review² that dealt in a thorough and detailed manner with the problem of the incorporation of satellite data (especially remotely sounded temperature data) into numerical prediction models but which omitted to describe some of the most recent experience at operational forecasting centres and which did not describe in any detail some of the plans for the future. Furthermore, it only dealt with satellite data and numerical models, and did not address (nor set out to) the role of satellite imagery in weather forecasting. Thus, your statement that, apart from temperature, satellites have little to say that bears on the meteorology of weather forecasting is incorrect. The skill achieved in modern operational forecasting is in no small part due to satellite information.

Let me mention two areas other than temperature where satellite observations are particularly important. First, winds at two levels in tropical regions are determined from the tracking of motion of clouds observed from geostationary satellites. This wind information is an important input into global models. Studies in which forecasts made with and without input from satellite data are compared demonstrate the value of satellite data in improving the performance of numerical models. For instance, from runs of the

numerical model at the European Centre for Medium Range Weather Forecasts at Reading Bengtsson³ concludes that satellite data enable the range of good forecasts to be extended from about 1.5 to 3.5 days in the Southern Hemisphere and (at least in some weather situations) from about 3.5 to 4.5 days in the Northern Hemisphere—a very useful improvement.

Second, satellite imagery, both visible and infrared, as received from geostationary and polar orbiting satellites provides an essential tool for operational meteorology. Cloud structure in data-sparse areas is interpreted in terms of synoptic patterns and appropriate data fed into models. Further, the up-to-the-minute information from such images, as interpreted by a forecaster, is essential for providing accurate short-range forecasts. The merging of satellite and radar data covering the United Kingdom and the surrounding region as in the FRONTIERS project⁴ is especially valuable.

For the future, what is required is first that the satellite observing system for operational meteorology consisting of five geostationary satellites and two polar orbiting satellites be made more reliable and robust. An important instrumental development here is that of the advanced microwave sounding unit which will provide better global temperature and humidity information from the polar orbiting satellites from the early 1990s; the UK Meteorological Office is providing the high frequency channels of this unit. A second requirement is that data from instruments such as the wind scatterometer on the European Space Agency's satellite ERS1 be properly exploited in operational meteorological models. Third, appraisal must be carried out of the

potential of new instruments such as the Doppler Lidar wind sounder which are candidates for flying on the polar platform for meteorological observation.

JOHN T. HOUGHTON

*Meteorological Office, London Road,
Bracknell, Berkshire RG12 2SZ, UK*

1. Maddox, J. *Nature* 325, 295 (1987).
2. Isaacs, R.D. *et al. Rev. Geophys.* 24 701–743 (1986).
3. Bengtsson, L. in *The Results of the Global Weather Experiment* (World Meteorological Organisation Publication, Geneva WMO 610, 1983).
4. Browning, K.A. & Carpenter, E.M. *Met. Mag.* 113, 282–288 (1984).

Oversight corrected

SIR—In our letter “Quantization effects in the photoelectrochemistry of superlattice electrodes”¹ we assumed that the observed photocurrent was generated only in the superlattice layers, and we completely neglected any contributions from the underlying GaAs_{1-x}P_x buffer layers or from a possible p–n junction adjacent to the p⁺–GaAs substrate. This assumption is valid for cathodic photocurrent in our particular sample configuration if the buffer layers are p-type; Mott–Schottky data from our strained-layer superlattice electrodes indicated p-type character. However, recent experiments show that the non-deliberately doped buffer layers show n-type behaviour, and that they contribute cathodic photocurrent at wavelengths above 775 nm. An additional and completely fortuitous complication is that the band gaps of the five buffer layers in our samples closely match five of the theoretical energy transitions for the GaAs quantum wells in our superlattices. These quantum wells comprise a total thickness of 0.5 μm and absorb 40–60 per cent of the incident light between 850–700 nm, respectively; the rest of the light passes into the buffer layer regions. Therefore, our observations and conclusions in this paper are clouded by the unrecognized contributions of the buffer layers.

Further work is now being carried out to separate the effects of buffer layers in strained-layer superlattice electrode structures from the effects of the superlattice only. We note that such experiments in lattice-matched superlattice electrodes, where graded-gap buffer layers are not present, do indeed reflect the quantization effects arising from the superlattice, and the structure of the photocurrent spectra agrees with the theoretical energy-level scheme of the GaAs quantum wells. But, the question of the occurrence of electron transfer from high-lying energy states in the quantum wells of superlattice electrodes into solution is now open. Work is in progress to answer the question.

A.J. NOZIK

*Solar Energy Research Institute,
Golden, Colorado 80401, USA*

1. Nozik, A.J., Thacker, B.R., Turner, J.A. & Olson, J.M. *Nature* 316, 51–53 (1985).

The worst of the worst

John Maddox

The Future of Immortality: And Other Essays for a Nuclear Age. By Robert Jay Lifton. Basic Books:1987. Pp.305. \$21.95.

ROBERT JAY LIFTON is an analytically inclined psychiatrist who, while a professor at Yale University, became renowned for his psychoanalytical questioning of people caught up in two recent tragedies — Hiroshima and the US war in Vietnam during the 1960s. More recently, Lifton has applied the same technique to a study of the conduct of the German physicians who served in Nazi concentration camps between 1939 and 1945.

In each case, his objective was to throw light of some kind on people's reaction to death in various manifestations. Why were the survivors of Hiroshima guilt-ridden above all? Why were the combatants in Vietnam both fearful for their own skins and callous about those of others (and guilt-ridden too)? And by what means could the camp physicians carry out their murderous duties behind a façade of personal normalcy?

These are all absorbing issues whose contemporary importance Lifton has never sought to hide. On the contrary, he makes no bones of his conviction that his findings bear directly on what he considers to be the psychopathology of the contemporary world — Armageddonism as he calls it. Lifton has been a prominent member of the US chapter of International Physicians for the Prevention of Nuclear War, as he was also engaged in what may be called the struggle against the Vietnam War.

Why such preoccupation with these gloomy themes? Lifton says that this latest book, *The Future of Immortality*, is in part an attempt to answer that question, which it does successfully. The content is a well-chosen collection of previously published essays (with one, on the psychoanalytical consequences of the Strategic Defense Initiative and related matters, specially written for this volume).

The clue is in the title: while individual immortality as taught to Christians is impossible, both Jung and Adler (but not Freud) give analytical weight to people's need to project their influence beyond their deaths (in which respect Lifton may be unfair to the Christian St Augustine, who did just that in his *City of God*). That is why people bother to make wills, or worry about their grandchildren's education.

Lifton calls this emotional projection beyond one's lifespan (backwards as well as forwards) connectedness. In his view, the future of immortality in that sense is

far from bright, threatened as it is by the Armageddonists.

Put briefly, Lifton's argument is that modern weaponry does psychoanalytical violence to people's sense of immortality, as he defines it. Hiroshima gave even the survivors a sense that the world had ended; by what twist of logic were they then able still to limp about upon the surface of the Earth? The prospect of an annihilating nuclear war (which phrase Lifton steadfastly puts in quotation marks to emphasize that there would be neither winners nor losers in the usual sense) must be similarly unsettling of people's emotional foundations. And so too must be the image of nuclear winter, which Lifton offers as the ultimate irony. Within his framework, he is no doubt right to argue that recent ameliorating calculations of nuclear winter (acknowledged in a footnote) are irrelevant; what matters is the image in people's minds.

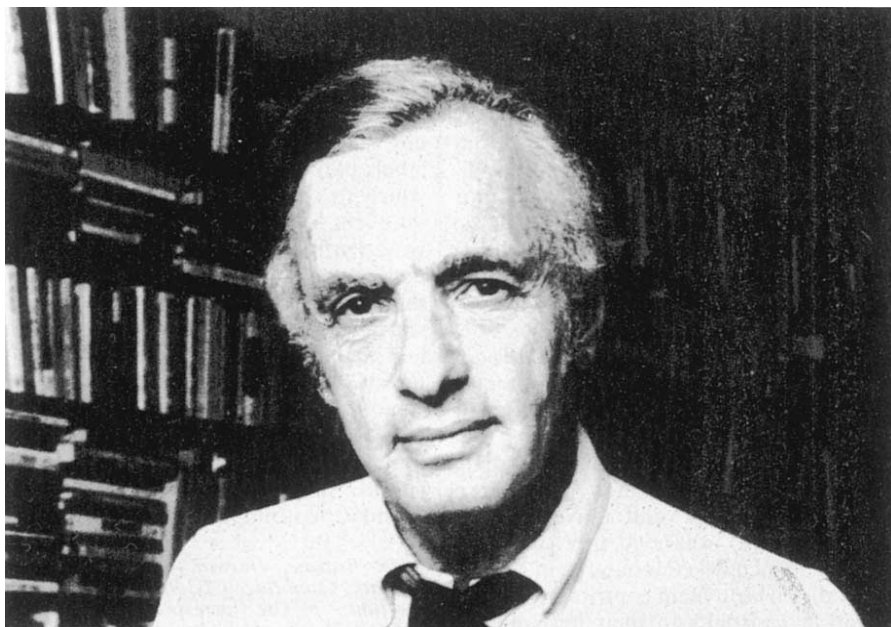
The first thing to say about this line of argument is that it should not be scoffed at. Lifton's first investigation, of the survivors of Hiroshima, was an inspired notion perhaps all the more revealing for having been carried out 15 years after the event, when the unfortunate had had time to reflect on their experience. Moreover, the fashionable scorn for psychoanalytical concepts notwithstanding, there is every reason why Lifton's argument deserves a sympathetic hearing, even by those quite

out of sympathy with his views on strategic issues.

Would it not be valuable to have some measurable (if qualitative) pointers to the reasons why nuclear technology, civil as well as military, engenders more alarm than other technology posing more immediate threats to individual survival, road traffic, for example? But Lifton could do his own case a power of good by breaking out of his chosen field of the psychoanalysis of cataclysm. If it is true that the image of the end of the world is inimical to a necessary concept of immortality (as defined), might not he expect to find a reinforcement of it among the lobbyists for and practitioners of space travel, for example? Unfortunately the only reference to this subject is to last year's shuttle accident, offered as evidence of the image of the "radical fallibility" of high-technology that Lifton says is now prevalent. (Chernobyl came too late to rate more than a footnote and a paragraph in the preface.)

Lifton's problem, in what is plainly as much a volume intended to influence opinion as to popularize a body of scientific work, is that his unyielding distaste for what he calls the Armageddonists prevents him from regarding them as also susceptible to his technique. "How can specific individual strategists and policy makers continue the nuclear buildup in the face of their knowledge of nuclear winter truths and Star Wars falsehoods?" He offers without direct investigation the explanation that those concerned are led to what he considers their error by a mechanism identical with that afflicting the Nazi doctors, while insisting that he does not "equate" the two groups.

For what it is worth, Lifton's demonology of the nuclear age is a predictable list with the exception of the inclusion of



Robert Jay Lifton — strength of commitment to the anti-Armageddonist cause.

Robert Oppenheimer, whose late recruitment to the ranks of the sceptical earns hardly any brownie points in Lifton's book. They are the ones who seek a sense of immortality by pretending that, even if nuclear wars cannot be won, it may be possible by cleverness to survive them (by civil defence, the Strategic Defense Initiative and so on). Lifton has no compassion for them.

This is where Lifton's collection of essays, while superbly written and thus compelling in a way, falls down. Psychoanalysts are perhaps best distinguished by their tolerance even of wayward behaviour born of their broad acquaintance with the human condition and their recognition

of the diversity of its causes. Lifton rises above this often pallid-seeming dispassion by the strength of his commitment to the anti-Armageddonist cause. At the same time, he is not above squabbles by mutual misquotation with others who have written more conventionally about nuclear strategy. The result is both a gloomy book and one with a curious blend of self-certainty and meanness that may do less than intended for the cause of anti-Armageddonism. But it is to be hoped that that will not deter others of a psychoanalytical inclination from more objective pursuits of similar goals. □

John Maddox is Editor of Nature.

Female attention

Janet Browne

Women in Science: Antiquity through the Nineteenth Century. By Marilyn Bailey Ogilvie. MIT Press: 1986. Pp.254. \$25, £19.95.

Women Scientists from Antiquity to the Present: An Index. By Caroline L. Herzenberg. Locust Hill Press, West Cornwall, Connecticut, USA: 1986. Pp.200. \$30.

THE women's movement has produced many books and the news of two more will not, perhaps, be of great consequence to many people. Yet it is worth pausing over these volumes because they attempt to do something rather different with the history of science.

The aim of both authors is to provide a roll-call of women from antiquity to the present century who were active in science, medicine and natural philosophy which will also convey an idea of the ebb and flow of science itself. It is not an 'alternative' view of history they wish to present but what you might call a 'realist' one, where midwives, physicians, astronomers and mathematicians were drawn from both sexes, but where women were usually outnumbered by men. Neither author braves the question of why this should be so, for they are not traditional feminists eager to ascribe the onset of the Dark Ages to male domination.

Marilyn Bailey Ogilvie describes only 180 women in order to give herself room to offer extensive biographical treatments and a long list of references. Also included is a short, judicious introduction that will serve its expressed purpose of sketching out the history of women in Western science for college and university students. The net was evidently cast wide enough to include a fair number of Victorian ladies whose main contributions were made as assistants to their husbands or fathers, but who nevertheless occupy a

precisely defined role in the advancement of science — one aspect that is often neglected. For example, it is good to see Mary Lyell, one of the five highly intelligent daughters of Leonard Horner and the wife of Charles Lyell, the geologist, listed in her own right as an accomplished conchologist. Others, such as Florence Nightingale, were evidently considered too well known to be included. The book consequently finds its limits in an unfocused though engaging way that will not always fully answer a reader's needs.

Caroline L. Herzenberg's book is much more to the point in this respect, being a checklist of 2,500 women with one or two words to give their professional interests and a bibliographic key letter to identify further sources of information. Most of the women one might think of seem to be here, although some sections such as midwifery and nursing could surely have been fleshed out by a week or two in a good library specializing in the history of medicine. The emphasis is strongly on the nineteenth and twentieth centuries, with a corresponding dependence on sources such as the *American Dictionary of Notable Women*.

Like Dr Ogilvie, Caroline Herzenberg deals mainly with Western science. Unlike her, she adopts an unfortunate system of labels derived from modern science with which to characterize the work of early scholars, so that Aristotle's esteemed wife, Pythias, is said to be "an important marine zoologist". This hardly represents the depth and sophistication of Pythias's views, nor does it grapple with the problem that she probably pursued these interests because her husband also did. Instead we are invited to believe that women have always followed their own inclinations, independent of their surrounding culture and society. Attractive as it sounds, this is simply not history. □

Janet Browne, Darwin Archives, University Library, Cambridge CB3 9DR, UK, is Associate Editor of The Correspondence of Charles Darwin, Vol. II of which was recently published by Cambridge University Press.

Moving back to the towns

Peter Warren

The Pyramid Builders of Ancient Egypt: A Modern Investigation of Pharaoh's Work-force. By Rosalie David. Routledge & Kegan Paul: 1986. Pp. 269. £16.95, \$37.50.

IN THE range and quality of its technology Pharaonic Egypt was pre-eminent among early civilizations. Our knowledge of this achievement is in part a function of the excellent preservation of artefacts, from pyramids to complex examples of carpentry and exquisite jewellery, but it comes chiefly from study of how the artefacts were made and from analysis of their components. Research of this sort has been carried out for many years. Today, however, we are at the beginning of a new generation of archaeological techniques. Powerful microscopes enable the fresh identification of metal and wooden objects, and allow reassessment of the processes by which they were manufactured; neutron activation programmes are transforming the analysis of fired clay, and thus having a great influence on studies into the provenance and trade of pottery; and the application of the techniques of organic chemistry is revealing the components of dyestuffs. The Egyptian database for all such investigations is immensely rich.

The book under review is concerned with some of these technological possibilities. Despite the title, only a small part of it deals directly with pyramid building, namely that at Lahun (and a third of this account is devoted to Middle Kingdom jewellery). The central subject is the towns of the royal workmen — conscript peasants, professional craftsmen and architects — and the way in which they lived.

These towns mark a specific form of urbanism in Egypt. They were deliberate creations, carefully located, planned within a thick brick wall, and guarded. Three of them are known in some details and are described in the book. At Tell el-Amarna the function of the workforce was to build and decorate the capital city and tombs of Akhenaten and his courtiers (14th century BC); at Deir el-Medina, near Thebes, to construct the royal tombs in the 'Valley of the Kings' and 'Valley of the Queens' (16th–12th centuries); and at Kahun to build the pyramid and associated temples of Sesotris II at nearby Lahun, and to maintain them thereafter (19th–17th centuries). The book gives no figures for the areas of the towns, there are no plans of Amarna or Deir el-Medina and that of Kahun has no scale. But it seems that

Kahun was large, perhaps occupying 16 hectares, and as such was about 24 times larger than Deir el-Medina. Yet Kahun is said to have had 100 houses against 140 in Deir el-Medina. (Data on pp. 59, 65, 102–105.)

The artefactual information from these two towns is supplemented by fragments of inscribed papyri, and Dr David has constructed a remarkably informative and detailed picture of the organization of the workforces, their working conditions, legal procedures, and systems of education, religion and medical care (especially gynaecological conditions and fertility tests). There is also insight into material life (a sickle for a left-handed person, a doll-maker's house with its stock of dolls' hair), as well as into manufacturing processes in the ordinary and the grander houses.

The latter part of the book describes a project at the Manchester Museum which is concerned with re-analysis of the Kahun assemblage. The chapter on textiles is full of detail, and that on neutron activation analysis (NAA) of metals is informative and suggests the existence of a local bronze industry (a table of results of the analyses would, however, have been a welcome inclusion). Ceramic analyses, also by NAA, are given a chapter, but much of this is background information, no quantified results are given and further work on Nile alluvial clays is said to be in hand. Similarly, analysis of the wood used at Kahun is promised in a future study. One wonders, therefore, whether the whole volume is not a little premature. It might have been better to have delayed publication until the interdisciplinary projects were complete.

There are also many points which specialists may question. For example, it would have been helpful to point out that the sycamore mentioned on p.66 is not the well-known *Acer pseudoplatanus* but the sycamore fig (*Ficus sycomorus*). The tomb of Maket, dated as 19th–20th dynasties on p. 108, is in fact correctly dated as 18th dynasty on p. 139. The summary of Egypto-Aegean relations (p176) is out of date. Kahun is not an "Early Bronze Age" town as stated on p.215, and it is surely not the case that "much of the pottery was clearly Aegean in style".

Editorially, things are not all they should be: only one of the four maps and plans has a scale and illustrations referred to on p.207 are missing, while text references to the plates and *vice versa* would have been helpful. But although this is a book marred by uneven presentation, and some inconsistencies and errors, it contains much to interest and inform both the expert and layman. □

Peter Warren is Professor of Ancient History and Classical Archaeology, University of Bristol, 11 Woodland Road, Bristol BS8 1TB, UK.



*Still keeping watch — seated statues of King Ramses II in front of the Great Temple at Abu Simbel, Egypt, painted by David Roberts in the mid-nineteenth century. The picture appears in *Temples and Tombs of Ancient Nubia: The International Rescue Campaign at Abu Simbel, Philae and Other Sites*, which was recently published by Thames and Hudson in collaboration with UNESCO, under the general editorship of Torgny Säve-Söderbergh. Price is £25.*

Enigmatic engineer

Norton Wise

Sadi Carnot: Reflexions on the Motive Power of Fire. Edited by Robert Fox. Lilian Barber Press, New York/Manchester University Press:1986. Pp.230. £29.95. US price not available.

SADI Carnot's classic memoir on the source of work from heat engines created no stir on its publication in 1824, nor for 20 years or so thereafter. Yet it argued in a most elegant and general manner that the work obtainable is independent of the working substance (steam, air and so on), depending only on the temperatures between which the engine operates, and that no work is obtainable without attendant losses of apparently available work. Coupling those ideas with conservation of energy, Rudolph Clausius and William Thomson fashioned the famous Second Law of Thermodynamics in 1850–1851. Since then, Carnot's work has inspired admiration, but also puzzlement. What were his sources; and why did his 'genius' go unrecognized?

In his critical edition, Robert Fox has supplied in English translation the materials upon which answers to these questions must be built: the text, Carnot's notes on related subjects, and an extensive commentary on the historical context and secondary literature. Strangely, from all of this apparatus we do not yet learn how Carnot came to his crucial view of work as deriving from the fall of heat from

high to low temperature, like the fall of water which drives a waterwheel; nor do we find a fully satisfactory answer to Carnot's failure to find an audience. Fox treats his materials with respect and reserve. Having striven for some years to unmask the real Carnot, he recognizes the pitfalls of hasty conclusions. He gives us the present state of his own research and that of others, criticizing, supporting and weighing as he sees fit.

Fox's main accomplishments are three. First, he sets Carnot firmly in the context of steam engineering in the midst of the industrial revolution, showing how Carnot shared the vision and many of the practical concerns of other engineers, but also why Carnot's abstract approach would have seemed foreign to them. Second, he makes good sense of Carnot's confusing series of results on the physics of gases, which occupy a large portion of the original memoir. Third, he draws out Carnot's increasing dissatisfaction with his own theory, which supposed heat to be conserved, and shows how this dissatisfaction probably affected the development of the memoir, ultimately leading Carnot to reject the caloric theory of heat and to adopt the view that heat is motion.

This critical edition will clearly be of great value to scholars, but it will leave them with a frustration. To use it they will have to collate the translation with Fox's earlier French edition. A more helpful volume would have included Carnot's original text. □

Norton Wise is a Professor in the Department of History, University of California at Los Angeles, Los Angeles, California 90024, USA.

Cardiac science in flower

Peter Harris

The Heart and Cardiovascular System: Scientific Foundations. Edited by Harry A. Fozzard, Edgar Haber, Robert B. Jennings, Arnold M. Katz and Howard E. Morgan. Raven: 1986. Two volumes, pp.1,797. \$225.

A YOUNG colleague tells me how, during a ritual post-meeting dinner, the distinguished bioscientist opposite lent over the table and offered his advice: "You should start looking into the metabolism of the heart, you know. There should be something there".

I wish I could afford to send him this magnificent two-volume overview of that now extensive field. The editors (any publisher's dream five-a-side) have gathered together contributions by 126 experts. Multiple chapters cover the ultrastructure, metabolic regulation, electrophysiology and contractile function of the heart. The structure and function of membrane systems, transport mechanisms and receptors are discussed in detail. There are excellent accounts of cardiogenesis, hypertrophy and ischaemia.

Publishing delays seem to have been minimized, so that there are references to work that appeared as recently as 1986. In this way the text will be useful as a reference source, but the book is also something to dip into with enjoyment. For somebody interested in the contractile apparatus, for instance, there are contributions on contraction, relaxation, contractile proteins, actomyosin ATPase, Starling's Law, ventricular mechanics, the mechanics of hypertrophied myocardium and a section of four chapters on excitation-contraction coupling. Somebody interested in electrophysiology has the choice of a general survey of the subject, chapters on membrane structure and membrane transport, the sodium pump, sodium-calcium exchange, ion channels, calcium and sodium channels, intracellular inorganic ions, gap junctions, transmembrane potentials and time-dependent outward currents, as well as seven chapters on arrhythmias.

A great deal of emphasis is given to methods, which are often given in sufficient detail to make the books useful on the bench. Here you will find technical information on the structural analysis of proteins, receptor binding, monoclonal antibodies, synthesis of peptides, nuclear magnetic resonance, the preparation of isolated myocytes, subcellular fractionation, lipid analysis and the use of high-performance liquid chromatography in the measurement of phosphometabolites. In addition there is a thoughtful contribu-

tion on experimental design in cardiac experiments. A further group of five chapters deals with imaging techniques. Some of these might be judged more appropriate to a clinical textbook, and the fine chapter on radionuclide imaging confirmed my suspicions that positron emission tomography has not, so far, provided the "important information concerning myocardial metabolism" suggested earlier in the book.

Although the title includes the cardiovascular system, the main emphasis is on the heart. The general microcirculation is little discussed, nor is the mechanics of blood-flow or blood-vessels, and the only regional circulation covered is the coronary. There is, however, a useful general review of smooth muscle.

These volumes look forward rather than backward. There is no attempt to duplicate standard textbook descriptions of metabolic pathways. More surprising is the omission of any discussion of

mitochondrial respiratory function. The great mass of human cardiovascular physiology which followed the introduction of the cardiac catheter is also almost entirely set aside. Instead the emphasis is on topical problems and on those techniques with which we are reaching into the future.

The organization of so many chapters has posed problems and their disposition in the different sections is sometimes capricious. The account of coronary circulation, for instance, is found in the section on cellular aspects of cardiac function. The reader will need, therefore, to use the index more than is usual.

These volumes bear eloquent witness to the immense and diverse flowering of cardiac science. I am glad of the opportunity to draw them to the attention of the general scientific public. □

Peter Harris is a Professor in the Cardiothoracic Institute, University of London, 2 Beaumont Street, London W1N 2DX, UK.

From up above

Robert N. Colwell

Remote Sensing: Methods and Applications. By R. Michael Hord. Wiley: 1986. Pp.363. £43.25, \$44.95.

THIS excellent book is yet more evidence that digital remote sensing has truly come of age in recent years. The author himself has wide experience in the subject. But by drawing heavily on a wide range of papers, reports and government documents, he has ensured both the breadth and the authoritativeness of his material.

The first eight tables have been compiled from NASA sources, and collectively they provide what the author terms "an overview of remote sensing". Thus, Table 1 compares wavelength regions by detectable parameters and constraints. Tables 2 to 6 systematically describe the areas of application and observable parameters for passive ultraviolet, passive visible, passive infrared, passive microwave and active microwave sensors. Table 7 catalogues observational needs with respect to each of nearly 100 parameters that are of interest to those concerned with the inventory, monitoring and management of the Earth's natural resources. And Table 8 systematically lists the many planned satellites for observing the Earth, indicating for each its objectives and orbital parameters together with the spectral responses and other relevant attributes of its sensors.

The remaining material is organized into three chapters. They deal, respectively, with sensors, processing and analysis techniques, and applications, the emphasis throughout being upon civilian rather

than military remote sensing. The first provides in-depth coverage of digital sensors, including radar imaging devices, coastal-zone colour scanners, thematic mappers and return-beam vidicon equipment. The next describes methods for converting remote-sensing-derived bits into pictures, identifying features from their resulting spatial/spectral recognition patterns, and evaluating the results. And the last gives an account of applications in the four main fields considered by the author to be of greatest economic importance: meteorology, geology, agriculture and oceanography. An Appendix lists the Band 5 pixel values for a portion of a Landsat Multispectral Scanner scene, included by Hord "so that the reader can try some digital processing on real data".

The book includes so little about cameras, films and photointerpreters that it might better have been entitled *Digital Remote Sensing*. Critics will have other complaints — the absence of colour figures; the overemphasis on figures from radar geology; the relative snubbing of applications in forestry and range management; the extreme terseness of some figure captions ("Schematic histogram", "Flow diagram", "State of Nevada"); and the limited usefulness of 29 pages of digits (the Appendix), which lack ground truth and seemingly will permit only "unsupervised classification" practice.

But criticism is cheap and I, for one, consider this to be an extremely useful book. I will make frequent use of it, and have no doubt that its intended audience — remote sensing technicians and others who have had the equivalent of an introductory course — will do likewise. □

Robert N. Colwell is a Professor in the Department of Forestry and Conservation, University of California, Berkeley, California 94720, USA.

Physical mechanisms for biased galaxy formation

Avishai Dekel* & Martin J. Rees†

* Racah Institute of Physics, The Hebrew University, Jerusalem 91904, Israel

† Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, UK

Most of the gravitating matter in the Universe is 'dark'. Galaxies are grouped into clusters and superclusters, and it is important to know whether the dark matter is distributed the same way. There are many reasons for suspecting that baryons may have condensed more efficiently into galaxies in regions of above-average density. The overall matter distribution would then be smoother than one would infer on the assumption that galaxies trace mass. The data are incompatible with the hypothesis that the Universe has the critical density unless galaxy formation was biased in this sense.

THE naive assumption that galaxies trace mass enables one to estimate the total mass density and to interpret the large-scale structure of the Universe based on the observed distribution of galaxies. However, galaxy type is already known to be correlated with environment¹, so it would be surprising if galaxy formation were not significantly affected by environmental effects segregating the galaxies from the underlying mass. We describe the motivation for this suspicion and argue that a segregation of one sort or another is a natural outcome of almost every cosmogony; there seems to be no difficulty in reconciling the observations with a 'flat' (density parameter, $\Omega = 1$) Universe. We then point at observations that may distinguish between various possible mechanisms for biasing.

Motivation

Dark matter and 'missing' mass. Since the introduction of the 'missing mass' problem by Zwicky more than 50 years ago strong evidence has accumulated revealing a discrepancy by a factor ~ 10 between the luminous mass and the dynamically inferred mass². Mass is generally estimated by using a relation of the sort $M \sim v^2 R$, involving some observed velocity v and size R (for example the virial theorem). With the associated luminosity obtained by photometry one can deduce a mass-to-light ratio, M/L . Writing the universal mean luminosity density as $L = 2 \times 10^8 \mathcal{L}_2 h \text{ Mpc}^{-3}$ (where the Hubble constant is $H_0 \equiv 100 h \text{ km s}^{-1} \text{ Mpc}^{-1}$ and $\mathcal{L}_2 \approx 1.5\text{--}3$ in the blue³), the measured M/L , if it is the same everywhere, implies a mean density relative to the closure density of $\Omega \approx (M/L)/(1,350 \mathcal{L}_2^{-1} h)$. A typical M/L value for the stellar content of ordinary galaxies is in the range 1–10, whereas the mass estimates give $M/L \geq 100\text{--}200 h$ in galaxies (rotation or X-ray data), and $M/L \approx 200\text{--}500 h$ in groups and clusters of galaxies (virial analysis), corresponding to $\Omega \sim 0.1\text{--}0.3$. (Taking into account the differences in stellar populations and gas content in galaxies and clusters, the implied ratio of luminous mass to total mass is actually roughly constant at ~ 0.1 on all scales from 50 kpc to 5 Mpc⁴.) Thus, although the inferred dark matter (DM) overwhelms the luminous matter, it is still short by a factor ~ 5 from 'closing' the Universe, confirming the classical conclusion of Gott *et al.*⁵. If the matter were all baryonic, $\Omega \approx 0.1$ would be compatible with standard Big Bang nucleosynthesis⁶, which constrains Ωh^2 , if $h \leq 0.65$.

Contrary to the consensus on the observational side, there has been a development in theorists' attitudes: motivated by particle theories, we are much more willing now to invoke non-baryonic DM that can contribute $\Omega = 1$, and thus satisfy the 'flatness' requirement. The inflation model⁷, being so appealing in resolving certain cosmological problems, enhances the credibility of the prejudice in favour of $\Omega = 1$, introducing a second DM problem (now deserving the original term 'missing mass', as there is no firm dynamical evidence for it).

For the currently observable part of the Universe to be in

causal contact at some earlier time, as required by the isotropy of the microwave background, the Universe must have inflated exponentially by a factor $\geq F_{\min} = z_{\text{inf}}^{-1} H_{\text{inf}}/H_0 \sim 10^{25}\text{--}10^{30}$ (where H_{inf} is the Hubble constant at the inflation epoch, whose redshift z_{inf} is either z_{GUT} or z_{Planck}). The present value of the density parameter is given by the Friedmann equation

$$1 - \Omega - \Lambda/(3H_0^2) = -k/(R_0 H_0)^2 \quad (1)$$

where R_0 is the current radius of curvature. In most versions of inflation the exponential growth, once started, continues rapidly for many expansion timescales: it is likely to overshoot, hence stretching any small part of an initial chaotic hypersurface so that it becomes essentially flat over our present horizon scale ($R_0 H_0 \gg 1$). With a vanishing cosmological constant ($\Lambda = 0$) this implies $\Omega = 1$ to a very high precision (limited only by the fluctuation amplitude in the microwave background, $\sim 10^{-5}$). For inflation to yield $\Omega \sim 0.2$, the inflation factor would have to be 'just' $\sim F_{\min}$, making the present radius of curvature of the order of the Hubble radius ($R_0 H_0 \sim 1$). This demands a fine-tuning which one would prefer to avoid. Moreover, if $\Omega \neq 1$, our currently observable Universe would (in a conventional inflationary model) have to grow from a segment of the initial hypersurface with the special property that its curvature was uniform to one part in 10^5 , otherwise the microwave background quadrupole would be larger than observed. Our Universe could therefore not have inflated from a typical element of an initial chaotic hypersurface.

Can the Universe be flat? How could we then explain the dynamical evidence for $\Omega \sim 0.2$? If $\Omega = 1$, the rising trend of M/L must continue out to scales of $\sim 10 h^{-1} \text{ Mpc}$. The mass estimates on such scales are based either on modelling the infall of the Local Group into the Local Supercluster (LSC)⁸, or applying a cosmic virial theorem to pairs of galaxies or to cluster-galaxy pairs⁹. In a linear, spherical model for the LSC

$$\Omega \approx \delta^{-1.7} (3v/Hr)^{1.7} \quad (2)$$

where δ is the density enhancement within the Local Group radius r , and v is the infall velocity at r . The cosmic virial theorem provides a dependence

$$\Omega \propto \xi(r)^{-1} (v/r)^2 \quad (3)$$

where v is the mean pair-velocity and $\xi(r)$ is the two-point correlation function, defined so that the probability of finding a galaxy at a distance r from another galaxy is higher by a factor $1 + \xi(r)$ than it would be if galaxies were distributed at random. The results apparently suggest again $\Omega = 0.1\text{--}0.3$. But the crucial point is that they are obtained using the quantities corresponding to galaxies: their number overdensity δ_g or the galaxy-galaxy correlation function $\xi_g(r)$. If, indeed, the galaxies cluster more than the underlying matter, so that $\delta_g = f\delta$ and accordingly (in the linear approximation) $\xi_g(r) = f^2 \xi(r)$, the real value of Ω , as

obtained by (2) or (3), is larger by a factor $f^{1.7-f^2}$. The data would be compatible with $\Omega = 1$ if the degree of bias corresponded to $f \approx 2-3$.

In rich clusters (c) the corresponding bias is the ratio of the mean cosmic M/L and $M_c/L_c \approx 300 h$, namely $f \approx 4.5 \mathcal{L}_2^{-1}$ if $\Omega = 1$. For Abell clusters of richness 1, where $\delta_g \approx 180$ within the Abell radius, the mass overdensity is therefore only $\delta \approx 40 \mathcal{L}_2$.

Evacuating voids. The emerging picture based on recent data¹⁰ is that 'voids' $\sim 50 h^{-1}$ Mpc in diameter in the galaxy distribution are actually quite common. The typical number density of (bright) galaxies in these voids, according to a crude estimate, is $<10\%$ of the mean. Based on spherical-void models¹¹, such an underdensity in the mass distribution would correspond at recombination to $|\delta| \geq 10^{-2}$ if $\Omega \approx 1$, and $|\delta| \geq 5 \times 10^{-2}$ if $\Omega \approx 0.1$. This would be incompatible with the isotropy of the microwave background on angles 10^{-1}° in any of the cosmogonic models, unless post-recombination reionization has washed out the fluctuations. The large-scale N -body simulations demonstrate the difficulty: even in 'pancake' simulations¹²⁻¹⁴ such large regions have not been found with density less than 25% of the mean; they cannot be substantially evacuated dynamically by the present epoch, commonly defined by matching the correlation functions of the simulated mass and the observed galaxies. The situation is worse if $\Omega < 1$, when it is even harder to evacuate the voids.

What is the real mass density in the voids? Consider a 'toy' universe which consists of superclusters and voids, each of a uniform density both in the matter and in the number of galaxies. As 'typical' measures of the galaxy density in superclusters and voids we take $\delta_g = 2.5$ and $\delta_g = -0.9$, these being approximately the values appropriate to the LSC and the Boötes void respectively. If $\Omega = 1$, the real mass overdensity in the LSC must be $\delta \approx 0.85$, implying a biasing factor $f \sim 3$. Without assuming any specific physical model of how galaxy formation might be inhibited in voids, we have (based on the definition of δ and δ_g)

$$(\delta_g/\delta)_{\text{void}} = (\delta_g/\delta)_{\text{sc}} = f \quad (4)$$

so the mass deficit in the voids must be $\delta \approx -0.32$. (The fractional volume in the voids is then $\sim 73\%$ and the corresponding mass is $\sim 50\%$.) The obtained mass densities in superclusters and in voids are then both compatible with $|\delta| \approx 9 \times 10^{-4}$ at recombination ($z \approx 10^3$). If most of the mass is non-baryonic, this corresponds to $\delta T/T \approx 3.5 \times 10^{-5} (\Omega h^2)^{-1}$, which is compatible with the observed isotropy constraints unless Ωh^2 is much smaller than unity. An open universe with $\Omega \approx 0.2$ would be in trouble here: it would require no bias, $f \approx 1$, and therefore, by (4), a real deep mass underdensity of $\sim 10\%$ in the voids. The corresponding $\delta T/T \sim 5 \times 10^{-3}$ would be hard to reconcile with observations without substantial reionization.

The relative 'efficiency' of galaxy formation in an environment x can be quantified by $\varepsilon = (n_x/\rho_x)/(n/\rho)$. From the illustrative values obtained above for n_x/n and ρ_x/ρ , the efficiency in voids is $\varepsilon \approx 0.15$ and in superclusters $\varepsilon \approx 1.9$. In rich clusters the efficiency is higher, $\varepsilon = (L_c/M_c)/(L/M) \approx 4.5 \mathcal{L}_2^{-1}$.

Difficulties in cosmogonic models. Simulations of the two most popular models^{2,15}, in which the Universe is either dominated by 'cold' DM (CDM) or by massive neutrinos, have led to the conclusion that (if $\Omega \sim 1$) neither can reproduce the observed large-scale distribution of galaxies unless galaxy formation is biased. In either case the two-point correlation function for matter steepens in time, and, if galaxies trace mass, the stage of the simulation to be regarded as the present epoch can be determined by matching its logarithmic slope to the observed $\gamma = 1.8$ of galaxies. It turns out that this timing is inconsistent with other aspects of galaxy formation and clustering.

The CDM density fluctuations at the onset of matter dominance, and the subsequent hierarchical build-up of structure, are determined by the assumption of adiabaticity and scale-invariance of the fluctuations as generated during inflation, and

by the 'cold' nature of the DM, which prevents free-streaming damping. The correlation function steepens in time as steeper parts of the spectrum become nonlinear. Davis *et al.*¹⁶ found that the 'present' correlation length in the CDM is only $\approx 1 (\Omega h^2)^{-1}$ Mpc, too small if compared directly with the $r_0 \approx 5 h^{-1}$ Mpc observed for galaxies, unless $\Omega h \leq 0.2$. Thus the specific model of CDM with $\Omega = 1$ would require an enhanced clustering of the galaxies relative to the DM such that $\xi_g(r) = (5-20) \xi(r)$ (for $h = 0.5-1$), which is consistent with the values of f deduced above from more general considerations.

For neutrinos with a mass in the range 10-100 eV, free streaming washes out all the fluctuations below $\lambda_v \approx 13 (\Omega h^2)^{-1}$ Mpc, comparable with the horizon scale when $kT \sim m_\nu c^2$, so the first objects to form are supergalactic 'pancakes' which subsequently fragment 'top-down' to galaxies. The neutrino correlation function steepens in a short time during the pancake formation as more material accumulates in the pancakes and as sub-clustering develops. The slope requirement $\gamma \approx 1.8$ implies that the first collapse to pancakes must have occurred recently, at $z \sim 1$ if $\Omega \sim 1$ (refs 12, 13). This poses a timing difficulty for any 'dissipative pancake' model which assumes that galaxies are 'daughters' of pancakes, as various lines of evidence suggest that galaxies began to form at $z \geq 3$ (for example, the existence of high-redshift quasars and galaxies, the great ages of globular clusters, evolutionary models for galaxies, and the need to 'hide' newly formed galaxies). Also, the neutrino correlation length, by the time that $\gamma \approx 1.8$, has increased to $\sim 5 (\Omega h^2)^{-1}$ Mpc, which is large in comparison with r_0 unless $\Omega h \geq 1$. If the galaxies form only in the collapsed regions as one would naively assume in this model, these constraints become even tighter; $r_0 \approx 8 (\Omega h^2)^{-1}$. Thus the bias required here is of an opposite sense: the galaxies should somehow be less clustered than the neutrinos.

The foregoing considerations all argue for some kind of large-scale segregation between galaxies and the dominant mass. If $\Omega = 1$, as favoured by theory, most of the (non-baryonic) mass must hide in the voids to escape detection in clusters and superclusters. Big regions, as empty as the 'voids' appear to be, cannot be evacuated yet, especially if $\Omega < 1$. (An $\Omega \approx 0.2$ model without biasing would require special non-random-phase initial conditions—for instance, galaxies might form from entropy (isocurvature) fluctuations whose amplitude is suppressed in incipient voids.) Finally, neither of the two best-studies cosmogonic models can match the observed large-scale structure unless galaxy formation is biased.

Bias mechanisms in general

Biasing can arise in three ways.

(A) The voids may be filled with a uniform component of 'missing mass' of a different nature from the clustered DM.

(B) Even if there were only one important kind of DM, the baryons may be segregated from the non-baryonic DM on scales up to $\sim 30 h^{-1}$ Mpc, allowing galaxy formation only in certain regions. This could result from large-scale 'isothermal' fluctuations, from gas dissipation relative to a collisionless DM (for example, neutrinos), or from very energetic winds of blast waves pushing the gas over large distances.

(C) Less extravagant in energy is the possibility that the large-scale baryon distribution does trace the DM on scales $> 1 h^{-1}$ Mpc, but the efficiency with which baryons turn into luminous galaxies is sensitive to other environmental effects. The bias could be determined in each protogalaxy autonomously, for example by its binding energy or local density, or it may be a result of feedback from other galaxies. This influence may propagate by gas transport to limited distances, or by radiation or fast particles to larger distances. The result might be destructive, suppressing galaxy formation locally (causing 'under-clustering') or far away (causing 'over-clustering'); but it could also be constructive, enhancing galaxy formation in the neighbourhood of other galaxies (for example, explosions).

Table 1 Biasing mechanisms

	Mechanisms	Observational tests
Do baryons trace mass on scales of ≥ 10 Mpc?	No $\rightarrow$ Large-scale explosions	Low gas density in voids 'Bubbles' in the distribution of galaxies Intergalactic gas very hot
Yes $\downarrow$	Primordial n_b/n_γ fluctuations	Large-scale structure without large associated fluctuations in microwave background
Initial galactic-scale perturbations gaussian?	No $\rightarrow$ String-induced perturbations	Large-scale structure without large associated deviations from Hubble flow Small 'steps' in microwave background temperature Special properties as gravitational lenses: double images with no magnification
Yes $\downarrow$	Secondary fluctuations (for instance, in pancake or explosion models)	Distinctive evolution of object clustering at high redshifts ($z = 1-4$)
Feedback processes from first galaxies important?	No $\rightarrow$ 'Autonomous' biasing—high-amplitude peaks convert into luminous galaxies more efficiently than lower-amplitude peaks of the same mass because of: 1) expectation that strongly bound haloes are more clustered than overall dark matter distribution; 2) effect of cosmological time (more pregalactic enrichment with heavy elements or higher density, more efficient cooling at onset of collapse); 3) more efficient retention of gas in first protogalaxies	Galaxies of shallower potential wells (dwarf galaxies) less biased
Yes $\downarrow$	Non-local negative feedback (rather than local positive feedback)?	Baryons in voids primarily in diffuse gas
No $\rightarrow$	'Epidemic' models of galaxy formation	Galaxies without dark haloes exist
Yes $\downarrow$	Baryons impeded from accumulating in galactic-mass potential wells?	'Failed' galaxies of low surface brightness exist
No $\rightarrow$	Star-formation efficiency (or IMF) in later-forming galaxies modified because UV background destroys H_2 , puts gas on higher adiabat and so on (this, may provide a local positive feedback effect instead)	Barren haloes should exist Dwarf galaxies more biased
Yes $\downarrow$	Gas heated to 10^6 K	
Thermal pressure of gas heated to 10^6 K, or pressure of cosmic rays, or Lyman- α photons coupled to gas, inhibits collapse. Pressure gradients push baryons at ~ 100 km s $^{-1}$ relative to dark matter		

We shall be concerned mainly with how mechanisms in categories B and C might operate in specific cosmogonic models; Table 1 may help to clarify our discussion. But before embarking on this, we venture a few comments on 'type A' biasing.

A uniform component. The Universe may be dynamically dominated by 'ultra-hot' weakly interacting particles which do not cluster because they have velocities $> 10^3$ km s $^{-1}$. For example, the 'ultra-hot' DM may be non-baryonic and the clumped DM all baryonic. But if the 'ultra-hot' particles are relics of an early epoch, their mass would have always been dynamically dominant over the baryons, and would have inhibited gravitational clustering altogether¹⁷. This would also yield an unacceptably fast expression timescale during nucleosynthesis. A way around this difficulty involves supposing that these particles arise from non-radiative decay of heavy particles with lifetimes only slightly shorter than the age of the Universe^{18,19}. Assuming that these decay products (presumably neutrinos) were substantially lighter than their unstable parents, they would be very 'hot'. Galaxies (haloes) and clusters would have formed during the era of matter domination by the unstable particles, but later expanded or even became unbound as a result of the decay. An elaboration of this idea suggests that the Universe finally becomes dominated by a stable primordial CDM species²⁰, which helps explain the survival of structure on both small and large scales, but this requires certain *ad hoc* fine-tuning among various DM components. Several astrophysical considerations constrain the decay epoch to be $1 + z_D \sim 5$; upper limits arise from the isotropy of the microwave background²¹, from the requirement that the cores of rich clusters were not disrupted during the decay²², and from the observed gravitational lenses if they are produced by normal galaxies and clusters²³. A lower limit is provided by modelling the infall of the LSC²² but it is very model-dependent²⁴. The flat rotation curves of galaxies, on the other hand, seem to be in conflict with significant decay into relativistic particles²⁵; this may rule out the model.

The universe could be flat ($k=0$) with $\Omega < 1$ if a non-zero cosmological constant contributed to the curvature such that $\Omega + \Lambda/(3H_0^2) = 1$ (equation (1)). In some respects this idea resembles the alternative discussed above; but, contrariwise, the Λ -term is unimportant at early epochs and so it would not have had such a serious inhibiting effect on galaxy formation. It is found in N -body simulations¹⁶ that the large-scale structure in a flat CDM model with $\Lambda \neq 0$ is quite successful in reproducing the two- and three-point galaxy correlation functions and their peculiar velocities (with no further bias in the galaxy formation). It is also compatible with the isotropy of the microwave background²⁶. However, for the Λ contribution today to be comparable with the ordinary matter, the required fine-tuning is as *ad hoc* as the one we intended to avoid by adopting $\Omega = 1$.

Bias in gaussian fluctuations. An enhanced clustering of galaxies over the background matter could arise in a 'bottom-up' model if galaxies formed only from exceptionally high peaks of the density distribution smoothed on galactic scales; peaks with an overdensity δ above a threshold $\nu\sigma$, where $\sigma^2 \equiv \langle \delta^2 \rangle$ and $\nu > 1$. Kaiser²⁷ showed that if the local distribution function of δ has a steeply decreasing tail, like a gaussian, high peaks occur with enhanced probability in the crests rather than the troughs of a large-scale fluctuation mode, and hence display enhanced clustering over the density contrast of the underlying matter distribution (if the distribution function has a flat tail the effect is weaker and it might even be reversed). For a gaussian process, in the region where $\xi(r) \ll 1$, the enhanced correlation function of high ν peaks is approximated by²⁸

$$\xi_{\text{peaks}}(r) \approx \exp[(\nu^2/\sigma^2)\xi(r)] - 1 \quad (5)$$

which becomes $\xi_{\text{peaks}}(r) \approx (\nu/\sigma)^2 \xi(r)$ when $\xi_{\text{peaks}} \ll 1$. The crucial question is what astrophysical mechanism prevents lower- ν peaks from also turning into galaxies, thereby neutralizing the effect (for example, ref. 29).

Autonomous bias. The high- ν peaks would collapse earlier, and

have higher density at turnaround, than more typical fluctuations on a given mass scale. They may, for this reason alone, form galaxies more readily and thereby create the biasing. Collapse and fragmentation of gas within protogalaxies are highly sensitive to the ratio of the cooling time ($\propto \rho^{-1}$ for bremsstrahlung and recombination cooling) to the collapse time ($\propto \rho^{-1/2}$): these processes occur more rapidly when this ratio is small. Moreover, if the high- ν perturbations collapsed at $z \geq 10$, Compton scattering of the microwave background would guarantee efficient cooling; this process, whose efficiency is proportional to $(1+z)^4$, would be unimportant for lower- ν systems collapsing at late times.

Protogalaxies that formed from high- ν fluctuations would have higher velocity dispersions (and escape velocities) than later-forming systems of similar mass. This makes it easier for them to retain gas that might otherwise be expelled by, for instance, supernova-driven winds. Dekel and Silk³⁰ have argued that in a bottom-up model the 'normal' bright galaxies must originate from high-density peaks (2σ – 3σ) in the initial fluctuation field; typical ($\sim 1\sigma$) peaks either cannot make a luminous galaxy at all because the gas is too hot and too dilute to cool in time, or, if their virial velocity is $\leq 100 \text{ km s}^{-1}$, they become diffuse dwarf galaxies after losing a substantial fraction of their mass in supernova-driven winds following the first burst of star formation. This would lead to a selective bias, in which the bright galaxies are biased towards the clusters and superclusters whereas the dwarf galaxies do trace the mass, and should provide an observational clue to the real distribution of the DM. The evidence for such a segregation between the high and low surface-brightness galaxies is still inconclusive³¹, because the relevant 'failed galaxies' are expected to be very faint and of very low surface brightness.

Negative feedback. The formation of galaxies from lower- ν perturbations may be further impeded owing to some kind of negative feedback induced by the first galaxies, by suppressing the baryonic infall or by affecting the stellar initial mass function (IMF). If these galaxies were able to heat the intergalactic medium to $>10^6 \text{ K}$, the Jeans mass would then rise high enough to prevent further protogalaxies from fully collapsing. The medium would readily be photoionized by ultraviolet from the first objects, but this process does not in itself raise the temperature much above 10^4 K . Although the energy required to heat the gas to $\geq 10^6 \text{ K}$ is modest (a few hundred eV per baryon) there is no natural way of supplying this energy efficiently in a sufficiently uniform way: blast waves with post-shock temperatures $\sim 10^6 \text{ K}$ do not propagate fast enough to yield homogeneous heating, and if the blast waves did travel fast enough to cross a proto-void region they would generate temperatures $\geq 10^8 \text{ K}$ and thus require much more energy. It is possible that the first galaxies would supply $\sim 100 \text{ eV}$ per baryon in some form which diffused rapidly, but was none the less sufficiently well coupled to the intergalactic gas to exert an effective pressure and thereby raise the Jeans mass. Cosmic rays, or photons near the Lyman- α resonance, are among the possibilities that have been discussed³².

If the galaxies result from infall of gas into a potential well due to non-baryonic matter (as in the CDM cosmogony), then their formation would be inhibited if the intergalactic gas were given a streaming velocity $\geq 100 \text{ km s}^{-1}$ relative to the dark matter (even if the gas stayed cold). This could happen if the first galaxies created large-scale pressure gradients (by radiation or cosmic ray pressure). The required streaming velocities, comparable with the escape velocities from the halo potentials, are much less than would be needed to evacuate voids completely of baryons; the energy requirements are therefore not excessive.

Even if feedback effects cannot inhibit the collapse of galactic-mass gas clouds, they can perhaps influence the internal evolution of protogalaxies, especially the stellar IMF. Ultraviolet radiation is capable of affecting the IMF in protogalaxies after the redshift $z \sim 3$ corresponding to the (apparent) peak of

activity of quasars. Silk³³ suggested that first-generation protogalaxies fragmented efficiently by H_2 cooling into a 'normal' stellar population. The radiation fed into the intergalactic medium would photodissociate the H_2 molecules, and prevent further formation during protogalactic collapse, leaving Ly- α emission as the dominant cooling mechanism, which made the fragmentation less efficient and led to a preferential formation of massive stars. The latter would be highly disruptive because of supernova-driven winds, eliminating bright galaxies and leaving behind only diffuse 'failed galaxies' from the second generation. However, an anti-bias might arise instead, if the fragmentation via H_2 were so efficient that it led to a population dominated by unseen 'Jupiters', while the later inefficient fragmentation ended up with a 'normal' visible population. Even the sign of this particular feedback effect will remain debatable until star formation is better understood.

A non-trivial constraint on any mechanism involving negative feedback is perhaps worth mentioning. Obviously the influence of the first bound systems must propagate fast. The time-lag between the turnaround of bound protogalaxies in an incipient cluster (where $\delta_{\text{cluster}} > 0$) and the turnaround of an equivalent fraction in an incipient void (where $\delta_{\text{void}} < 0$) is $\Delta t/t \approx \delta_{\text{cluster}} + |\delta_{\text{void}}|$, where t is the cosmological timescale and both δ_{cluster} and $|\delta_{\text{void}}|$ are still < 1 ; unless the influence pervades the void within a time Δt it plainly cannot pre-empt the formation of galaxies there. But it would seem even easier for a galaxy to quench formation of new neighbours—whatever the actual feedback mechanism was it would surely be stronger at closer range, and the energy released within an incipient cluster might then be inadequate to exert any feedback on larger scales. This difficulty would not arise if there were a delay t_{delay} between the onset of protogalactic collapse (defined as the latest stage at which negative feedback could stop the protogalaxy from evolving into a luminous galaxy) and the stage when it would first exert negative feedback. The first galaxies would then be unable to quench the formation of others (even close neighbours) unless they lagged behind the first ones by longer than t_{delay} . Remote negative feedback would still occur if the influence of the first galaxies reached the void in a shorter time than $\Delta t - t_{\text{delay}}$. The physical interpretation of t_{delay} depends on the actual feedback mechanism. If, for instance, this involved quasar-like activity in the first galaxies, t_{delay} would be the time taken for a young galaxy to develop violent activity in its nucleus. **On galaxies versus baryons.** The LSC may be used to test the general possibility that the baryons trace the mass and that the bias relies on the sensitivity of galaxy formation to the background density^{9,34}. Assume that when galaxies formed, at some z_i , their mean number density contrast within the proto-supercluster was enhanced by a factor g over the matter contrast, $\delta_g^i = g\delta^i$. Consequently, the linear growth of δ_g would be driven by the δ of the dominant mass, leading to a present bias of

$$\delta_g = f\delta \approx \frac{(g + z_i)}{(1 + z_i)} \delta \quad (6)$$

For f to be ~ 2.5 as required by $\Omega = 1$ (see equation (1)), the original bias, if $z_i \geq 3$, would have to be $g \geq 7$, which may look somewhat large. But actually it provides an interesting consistency check for such a bias if it is due to high gaussian peaks, and if the Universe has $\Omega \sim 1$ and is dominated by CDM. The bias of high peaks in a gaussian density field is a function of the r.m.s. mass density contrast at z_i on co-moving galactic scales, σ^i , and the height of the threshold, ν . Given that for galaxies today $\sigma_g \approx 1$ at $8 h^{-1} \text{ Mpc}$, one can write $\sigma^i = (1 + z_i)^{-1} s$, where $s \equiv \sigma^i / \sigma^i(8 h^{-1} \text{ Mpc})$, a parameter describing the steepness of the linear fluctuation spectrum. The amplification at z_i is given by²⁷

$$1 + \delta_g^i = \exp(\nu \delta^i / \sigma^i) = \exp(\nu \delta_g / s) \quad (7)$$

The initial g can be predicted from (7) by using $\delta^i = \delta_g f^{-1}(1 + z_i)^{-1}$, to be compared with the value deduced from

(6). The epoch of galaxy formation is determined by the requirement $\nu\sigma^i \approx 1$, that is, $(1+z_i) \approx s\nu f^{-1}$. Equating g in (6) and (7) we finally get

$$f^{-1} \approx 1 - \delta_g^{-1} [\exp(\nu\delta_g/s) - 1] + (\nu s)^{-1} \quad (8)$$

Note that the present bias f is insensitive to z_i but is quite sensitive to the shape of the spectrum through s . If we adopt for galaxies at formation a mass of $10^{11} h^{-1} M_\odot$, the CDM spectrum¹⁵, for $\Omega = 1$ and $h = 0.5$, gives $s \approx 6$. With $\delta_g = 2.5$ and $\nu = 2.5$ we get in (8) $f \approx 2.6$, a value in agreement with the degree of bias required in the first section. This pleasant consistency would not have occurred if the DM were not 'cold'; a steeper spectrum would give a weaker bias f (which may in turn be consistent with some $\Omega < 1$). For example, a white-noise power spectrum ($n=0$) would give only $f \approx 1.08$. If $s < 5$ the bias deduced from (6) is too low to match the amplification provided by (7). Turning the argument around, if we start by assuming CDM, this consistency would require $\Omega = 1$. This is because on one hand a smaller Ω would be associated with a smaller s in the CDM spectrum so f would be much larger in (8), whereas on the other hand a smaller Ω needs less 'missing' mass to be reconciled with the observed galaxy distribution: that is, a smaller f .

The above estimate of the required g ignores the effects of asphericity and nonlinearity on the dynamics of the LSC. For example, the high local density resulting from the one-dimensional collapse of pancakes may provide a mechanism efficient enough to yield the required bias.

Specific cosmogonic models

Biasing in the CDM cosmogony. The standard CDM cosmogonic scheme ($\Omega = 1$ and scale-invariant initial curvature fluctuations) has been investigated in greater detail than any other equally specific model. Detailed simulations¹⁶, hypothesizing that galaxies form only from regions which are $>2.5\sigma$ peaks of the initial mass distribution (smoothed on a galactic scale), show a good fit with the galaxy-galaxy correlation function, the M/L ratio in clusters and the Virgo infall. The resultant haloes also have profiles and velocity dispersions compatible with observation³⁵. This is encouraging enough to motivate us to consider whether any of the mechanisms discussed in the previous section can cause the requisite biasing in the specific context.

The CDM cosmogony will not be able to predict the luminosity function of galaxies until we develop a physical understanding of several distinct phases of galaxy formation.

(1) Baryons are presumed to condense in virialized 'haloes' of dark matter in the approximate mass range 10^8 – $10^{12} M_\odot$. For larger masses, dissipative cooling may be inefficient; below $\sim 10^8 M_\odot$ the potential wells may be too shallow to capture primordial gas (especially if this gas has been ionized and heated to $\sim 10^4$ K). The mass distribution of isolated virialized systems can in principle be learnt from N -body simulations. In practice, progress is impeded by the large dynamic range involved; galactic-mass haloes build up hierarchically, via a complicated series of mergers, between $z \approx 5$ and the present epoch. Present simulations (for example refs. 35, 36) have a mass range of $\sim 10^4$.

(2) Even if the dissipationless clustering of the dark matter were accurately known, the fate of the baryonic component—how much gas falls into each potential well, and how much is retained—involves complex gas dynamics.

(3) The resultant luminosity and brightness profile of the galaxy depends on the IMF of the stars that form from the gas, and on the efficiency of star formation.

The biasing could in principle involve any (or all) of these three aspects of galaxy formation. The N -body simulations³⁶ suggest that even stage (1) may lead to substantial biasing: the dark haloes with circular velocities $v_c > 200 \text{ km s}^{-1}$ —those that might be associated with bright galaxies—are themselves more clustered than the overall mass distribution, essentially because the binding energy of galactic mass perturbations is significantly

affected by whether they are in a peak or a trough of the fluctuation spectrum at larger wavelengths. (The linear growth rates are boosted or suppressed if they are, in effect, in a background where the overall expansion mimics an $\Omega > 1$ or an $\Omega < 1$ universe respectively.) This effect alone might account for the requisite biasing of bright galaxies (although it would not explain the tendency of galaxies of low luminosity to cluster like brighter galaxies (for example, ref. 10, based on Eder, Schombert, Oemler and Dekel, in preparation)). Nevertheless, any of the other mechanisms involving autonomous biasing or feedback could also be operative, modifying stages (2) and/or (3) of the formation process.

Although galaxy formation in the CDM cosmogony is a relatively recent cosmic phenomenon, occurring at $z \approx 4$, bound systems of subgalactic masses would appear somewhat earlier. The first such systems would have condensed at $z \approx 10$ – 20 from high- ν peaks in the density distribution on scales $10^6 M_\odot$. Even a fraction 10^{-4} of the cosmic baryons, forming massive stars within these first bound systems, could release enough energy to reionize all the matter, thereby raising the Jeans mass to $>10^8 M_\odot$ and quenching the collapse of further baryonic systems of $10^6 M_\odot$. The only $10^6 M_\odot$ objects that ever form may therefore arise from $>3.5\sigma$ peaks, and consequently display enhanced clustering. In particular, they will be clustered on galactic scales (10^{11} – $10^{12} M_\odot$) in such a way that there are more than twice as many in a 2.5σ galactic-scale peak as in a 1σ peak on the same scale. If this resulted in star formation proceeding differently in 'high' and 'typical' galactic-scale perturbations (by, for instance, producing a significant heavy element abundance, or triggering some runaway process in the former), this could provide another physical reason for the prescription that only the high- ν peaks evolve into galaxies: they could have developed internal differences even before turnaround³⁷.

Biasing in pancake models (for example neutrinos). A bias is generated automatically in any 'top-down' model where the perturbations below a coherence length $\sim 30 h^{-1} \text{ Mpc}$ have all damped out, as with adiabatic perturbations in a neutrino-dominated or baryonic universe. First, there are motions from 'proto-voids' to 'proto-pancakes' associated with the large coherence length; collapse into flat pancakes is accompanied by streaming toward their lines of intersections ('filaments'), and toward the knots where rich clusters form. The gas then concentrates dissipatively into the high-density regions, within which the conditions become ripe for cooling and galaxy formation; galaxies are thus limited to very specific regions. (This is essentially a combination of type B and type C mechanisms.)

However, if the efficiency of galaxy formation were similar in all the collapsed regions, this natural bias would lead to problems in understanding how 'pancakes' could have collapsed soon enough to form galaxies without producing large-scale clustering of excessive amplitudes. This problem could only be avoided if galaxy formation had somehow been suppressed in the high-density regions (or, less likely, enhanced in the low-density regions). For instance, galaxies might form preferentially in the flat sheet-like pancakes and not in the denser filaments and clusters, maybe because cooling was more efficient behind shocks of a planar geometry, or because galaxy formation was, for some reason, more efficient at later times, when most of the pancake galaxies formed. N -body simulations, in which the formation of a galaxy at a given position and time was determined taking into account suppressing feedback effects from nearby quasars³⁸, demonstrate that the required anti-bias could be easily obtained with a reasonable choice of values for the physical parameters such as the quasar output energy and lifetime and the cooling rate of the heated gas.

Although an anti-bias like this can eliminate the timing-scaling problem, one has to worry about reconciling it with the indications for a positive bias summarized earlier, and with the excessive X-ray emission predicted from the big neutrino clusters if the gas has accumulated in their cores³⁹. The solution to the

former may involve anti-bias on scales of clusters and bias on scales of superclusters and voids. The latter is avoided if the inter-cluster gas has been heated by shocks to higher adiabats before contraction into clusters⁴⁰.

An alternative resolution is non-dissipative pancake collapse⁴¹, as would arise from a hybrid model: if galaxies form independently of pancakes, from another component of density fluctuations, the timing constraint becomes irrelevant: galaxies could have formed at $z \geq 3$ and large-scale pancakes at $z \sim 1$. Galaxies would not be limited to pancakes but would rather be present everywhere, subject to the biasing mechanisms that are relevant in general 'bottom-up' models. Such hybrids could involve two types of DM, baryonic and/or non-baryonic; or else two types of initial fluctuations, adiabatic and isothermal^{14,42-46}. The hybrid models can be successful where the single-DM models fail: for example, in reproducing simultaneously the observed structure on galactic scales and on supergalactic scales, and in smearing the anisotropies in the microwave background.

Biasing in the explosion model. Biasing also results if galaxies enhance the formation of other galaxies close to them. In the 'explosions' model suggested by Ostriker and Cowie⁴⁷ and by Ikeuchi⁴⁸ such effects dominate the cosmogonic process. Here, nuclear energy from first-generation objects helps gravity in forming further galaxies, hence enhancing their clustering. The exploding galaxies produce spherical blast waves that push the gas out of their interiors. The shells expand, cool and fragment into a new generation of galaxies, the last generation forming at $z \approx 7$, after which the shells cannot cool efficiently any more. The galaxies hence form in spherical shells while cool gas still fills the inter-shell volume, and the non-dissipative DM is still distributed everywhere. This introduces an automatic bias in the galaxy positions similar to that in pancake models. The lifetime of this bias is limited: in addition to the accretion of DM onto galaxies, the evacuation of the baryons from the shell interiors creates under-dense regions which then tend to develop gravitationally into voids empty of DM. N -body simulations⁴⁹ with $\Omega = 1$ and $\Omega_b = 0.1$ show that a big enough bias ($f \geq 2.5$) can persist until a universal expansion by ~ 8 after the formation of the last generation of galaxies, which is compatible with the favoured epoch for galaxy formation in this model.

It is possible that the explosive shells under their mutual interactions are responsible for the observed large-scale structure (ref. 48; Weinberg, Dekel and Ostriker, in preparation). Even if they are not, it would be surprising if positive feedback effects ('epidemic' galaxy formation) did not play some role on the scale of individual groups and clusters: if the energy output from a galaxy could compress surrounding gas by a factor ≥ 10 , then this gas could undergo gravitational collapse unassisted by dark matter.

Very-large-scale structures

There are indications that the very-large-scale superclusters and voids may call for a formation (and bias) mechanism over and above the type of biasing process that could plausibly account for the galaxy correlations on scales $< 10 h^{-1}$ Mpc. One crucial observation in addition to the presence of big voids is the enhanced superclustering of rich Abell clusters over the galaxies (and the matter)⁵⁰⁻⁵², characterized by a correlation function nonlinear out to $\sim 25 h^{-1}$ Mpc and claimed also to be associated with peculiar velocities of $\sim 1,000 \text{ km s}^{-1}$. The cluster correlation function remains significantly positive out to $\sim 50 h^{-1}$ Mpc. On larger scales the data are less significant (ref. 53 and N. Kaiser, personal communication), but there is no evidence that the function goes negative, as would be expected in the CDM model ($\Omega = 1$ and scale-invariant initial spectrum). Another puzzling finding of potentially great importance^{54,55} is the apparent bulk motion of a region $\sim 100 h^{-1}$ Mpc in diameter around us relative to the microwave background frame of reference with a velocity of $\sim 600 \text{ km s}^{-1}$. This superstructure, on such large scales, is

hard to reproduce in any of the standard models that assume initial fluctuations with a scale-invariant spectrum and random phases⁵⁶. In particular, the standard CDM model predicts that the matter correlation function should go negative beyond $\sim 20 h^{-1}$ Mpc, which, if the clusters are biased as in equation (5), implies that the cluster correlation function should also become negative there, in possible conflict with the observations. The predicted r.m.s. bulk velocity on such scales is $\sim 150 \text{ km s}^{-1}$: well below the claimed value.

One way to explain such very-large-scale structure and still be consistent with the microwave background isotropy is by appealing to an open universe, where the standard assumption that the initial curvature fluctuations are scale-invariant leads to a post-recombination spectrum with more relative power on large scales. If $\Omega \sim 0.2$ (of which CDM contributes $\Omega \sim 0.1$), then baryons or neutrinos give rise to a positive feature in the spectrum at $\sim 100 h^{-1}$ Mpc while CDM can still be responsible for the formation of galaxies (see, for example, Dekel, Blumenthal and Primack, in preparation). But if $\Omega \sim 0.2$ no net bias is required by the dynamics of the LSC or the cosmic virial theorem, indicating in this case that we have overestimated the 'emptiness' or the extent of the big voids.

If the density fluctuations began as quantum fluctuations of a free scalar field during the era of inflation they are indeed expected to be gaussian⁵⁷, but it is also possible that the fluctuations arose from a different mechanism, in which they would not in general be gaussian, and have non-random phases. If the initial fluctuations had non-random phases, then the Universe might be inherently smoother in some places than in others: galactic-scale fluctuations with amplitudes large enough to form bound systems could then abound in some regions, but be entirely absent in large 'calm' patches, the latter corresponding to voids.

A specific model that incorporates this feature is one in which the density fluctuations were induced by cosmic strings⁵⁸. The strings are curvature singularities with a random-walk topology generated in a phase-transition in the early Universe (for example, associated with the gauge group $U(1)$), which turn into closed smoother 'parent' loops on entering the horizon and then chop themselves into (presumably) stable 'daughter' loops, all in a scale-free self-similar fashion. Based on the first, low-resolution string simulations⁵⁹, it was argued that the final loop-loop correlation function has a general shape close to that of galaxies or clusters⁶⁰, $\xi(r) \propto r^{-2}$, as expected from 'beads' along locally linear 'strings'. The string-induced fluctuations in the DM (by accretion onto the loops), with a scale-invariant Harrison-Zeldovich spectrum, would be the seeds for the formation of galaxies and clusters, which are therefore expected to be aligned in space along the same 'parent' linear structures. When we studied a toy string model that incorporates DM gravity⁶¹, these phase correlations were found to have a very pronounced effect on the correlation functions of the galaxies and the clusters relative to the matter without violating the isotropy of the microwave background. It provides a galaxy biasing mechanism, as well as a possible explanation for the excess of cluster clustering on very large scales.

Unfortunately, there are reasons to believe that this understanding of the clustering of string-loops is premature. First, it turns out that the correlations on scales larger than the horizon are negligible because the strings resemble a self-avoiding random walk (E. Vishniac, personal communication; Blumenthal, Dekel and Primack, in preparation). Second, it is not obvious that the notion of 'beads along strings' is at all relevant; new simulations⁶² show no such effect. Also, high loop velocities tend to smear out their correlations on scales slightly smaller than the original parent loops (on the same scale as the horizon). It is therefore essential that we go back to the basic ingredients of the string model and study the fragmentation process of strings and the associated velocities in more detail before a serious attempt is made to understand the formation of large

scale structure in this model.

Biasing mechanisms based on favouring high- σ peaks can enhance the amplitude of correlation functions, but cannot reverse their sign. In particular, they cannot, in the CDM model, yield positive correlations on scales exceeding $\sim 20 h^{-2}$ Mpc, where the mass-correlation function is negative (because the spectrum on large scales falls off more steeply than white noise). However, some feedback processes could, in principle, generate apparent large-scale positive correlations, as in the following example. We know that some galaxies at $z \approx 3$ develop powerful quasar-type activity in their nuclei; indeed, the ultraviolet or kinetic energy output from these objects is widely believed to have a dominant influence on the thermal history of the intergalactic medium. The active lifetime of individual quasars is unknown, but could be as long as $\sim 2 \times 10^9$ yr, the timescale for cosmological evolution at $z \approx 3$. A quasar emitting $> 10^{46}$ erg s $^{-1}$ for this time span would release enough energy to ionize all material within a sphere of radius $10 h^{-2/3} (\Omega_b/0.1)^{-1/3}$ Mpc (corresponding to $40 h^{-2/3} (\Omega_b/0.1)^{-1/3}$ Mpc when expanded to the present epoch); quasars could supply ~ 100 eV per baryon (sufficient to raise the Jeans mass to galactic scale) over a volume of present radius $\sim 20 h^{-2/3}$ Mpc, if $\Omega_b = 0.1$, and thereby quench the further growth of small-amplitude galactic perturbations. The energy supply would be too diffuse to halt the contraction of protogalaxies that were already undergoing nonlinear collapse; however if the ultraviolet were not attenuated by dust and did indeed modify star formation (compare ref. 33), it could still affect the luminosity or surface brightness of the resultant galaxy. The distribution of bright galaxies that form later might then be modulated on scales corresponding to the range of influence of such quasars; moreover, if the output from a quasar were directional, each object could imprint correlations on still larger scales than for the isotropic case. This is just an example (and models involving cosmic strings can provide others) to illustrate that the large-scale distribution of galaxies could be modulated by non-gravitational effects: it might then reveal very little about the overall mass distribution on those scales, and no non-Hubble velocities need then be associated with the apparent inhomogeneities.

Conclusions

A few key observations may be helpful in distinguishing between the possible bias mechanisms. In relation to the voids one would like to answer questions like:

(1) How big are the voids and how empty are they? We need to quantify the data by using a meaningful statistic, to confirm (or disprove) our assessment that none of our 'standard' models can account for the voids without bias.

(2) Are voids empty of galaxies of all types, or only those

types that are most conspicuous? Any evidence that galaxies of different types display unequal degrees of large-scale clustering would be relevant here, and most interesting would be the spatial distribution of very low-surface-brightness dwarfs.

(3) How much gas is there in the voids? Absorption systems along the lines of sight to quasars passing through voids may be detectable. If the gas is at 10^5 K, too hot for 21-cm radiation and too cold for X-rays, perhaps some features characteristic of neutral He may reveal its presence. The clustering of Ly- α clouds along the lines of sight to quasars which pass through voids may provide useful information.

The relationship between 'parent' haloes and 'daughter' galaxies could also have interesting implications on the bias schemes⁶³:

(4) Are there any 'barren' galactic-mass dark haloes with no luminous galaxy within them? Such objects may be associated either with small haloes (shallow potential wells) or with haloes too big to let the gas cool in a Hubble time. (They may, perhaps, be candidates for invisible gravitational lenses).

(5) Are there any galaxies which lack dark haloes? Such galaxies may form from regions where the baryons had been compressed by non-gravitational forces to ~ 10 times the DM density, indicating certain types of bias mechanisms (Table 1).

We have argued that biasing is not just an *ad hoc* contrivance introduced by theorists to save the philosophically attractive $\Omega = 1$ model when confronted with apparently conflicting evidence. The gross features of the large-scale structure, in particular the big voids and the superclustering of clusters, are strongly suggestive of biasing, and it is mandatory if we wish to reconcile any of the currently-favoured cosmogonic models with the observed Universe. Therefore the search for an appropriate physical bias mechanism, and especially for confirming observational evidence, is of great importance. What might have looked at first as a frustrating idea for astronomers, who can only observe the luminous 'tips' of the 'icebergs', should become an exciting area of observational search for the relevant circumstantial evidence.

On the theoretical side, finding the true bias mechanism is intimately related to elucidating the correct cosmogonic model and the nature of the DM. Although the proposed bias mechanisms are hard to quantify, they are very plausible physically: some of them, at least, must have affected galaxy formation to some degree and they should be worked out in more detail. It is evident that the notion that galaxies trace mass is an unjustified assumption.

We thank Carlos Frenk, Nick Kaiser, Adrian Melott and Jim Peebles for helpful comments, and gratefully acknowledge the hospitality of the University of California, Santa Cruz, where part of this paper was written.

- Dressler, A. *Astrophys. J.* **236**, 351-365 (1980).
- Dekel, A., Einasto, J. & Rees, M. J. *Rev. mod. Phys.* (in the press).
- Felten, J. E. *Comments Astrophys.* **11**, 53-67 (1985).
- Blumenthal, G. R., Faber, S. M., Primack, J. R. & Rees, M. J. *Nature* **311**, 517-525 (1984).
- Gott, J. R., Gunn, J. E., Schramm, D. N. & Tinsley, B. M. *Astrophys. J.* **194**, 543-553 (1974).
- Yang, J., Turner, M. S., Steigman, G., Schramm, D. N. & Olive, K. A. *Astrophys. J.* **281**, 493-511 (1984).
- Guth, A. *Phys. Rev. D* **23**, 347-356 (1981).
- Davis, M. & Peebles, P. J. E. *Astr. Astrophys.* **21**, 109-130 (1983).
- Peebles, P. J. E. *Nature* **321**, 27-33 (1986).
- Oemler, A. Jr in *Nearly Normal Galaxies* (ed. Faber, S.) (Springer, New York, in the press).
- Hoffman, G. L., Salpeter, E. E. & Wasserman, I. *Astrophys. J.* **263**, 485-500 (1982).
- White, S. D. M., Frenk, C. S. & Davis, M. *Astrophys. J.* **274**, L1-L5 (1983).
- Centrella, J. & Melott, A. *Nature* **305**, 196-198 (1983).
- Dekel, A. & Aarseth, S. J. *Astrophys. J.* **283**, 1-23 (1984).
- Primack, J. R. in *Varenna Lectures* (Academic, New York, 1986).
- Davis, M., Efstathiou, G., Frenk, C. S. & White, S. D. M. *Astrophys. J.* **292**, 371-394 (1985).
- Hoffman, Y. & Bludman, S. *Phys. Rev. Lett.* **52**, 2087-2089 (1984).
- Turner, M. S., Steigman, G. & Krauss, L. M. *Phys. Rev. Lett.* **52**, 2090-2094 (1984).
- Gelmini, G., Schramm, D. N. & Valle, J. W. F. *Phys. Lett. B* **146**, 311-317 (1984).
- Olive, K., Seckel, D. & Vishniac, E. *Astrophys. J.* **292**, 1-11 (1985).
- Silk, J. & Vittorio, N. *Astrophys. J.* (in the press).
- Efstathiou, *Mon. Not. R. astr. Soc.* **213**, 29P-34P (1985).
- Dekel, A. & Piran, T. *Astrophys. J.* (in the press).
- Hoffman, Y. in *Galaxy Distances and Deviations from Universal Expansion* (eds Madore, B. F. & Tully, R. B.) 237-242 (Reidel, Dordrecht, 1976).
- Flores, R., Blumenthal, G. R., Dekel, A. & Primack, J. R. *Nature* (in the press).
- Vittorio, N. & Silk, J. *Astrophys. J.* (in the press).
- Kaiser, N. *Astrophys. J.* **284**, L9-L12 (1984).
- Politzer, D. H. & Wise, M. B. *Astrophys. J.* **285**, L1-L3 (1984).
- Bardeen, J. in *Inner Space/Outer Space* (eds Kolb, E. W. & Turner, M. S.) 212-217 (University of Chicago Press, 1985).
- Dekel, A. & Silk, J. *Astrophys. J.* **303**, 39-55 (1986).
- Haynes, M. P. in *Nearly Normal Galaxies* (ed. Faber, S.) (Springer, New York, in the press).
- Rees, M. J. *Mon. Not. R. astr. Soc.* **213**, 75P-81P (1985).
- Silk, J. *Astrophys. J.* **297**, 1-8 (1985).
- Dekel, A. *Comments Astrophys.* **11**, 235-256 (1986).
- Frenk, C. S., White, S. D. M., Efstathiou, G. & Davis, M. *Nature* **317**, 595-597 (1986).
- White, S. D. M., Frenk, C. S., Davis, M. & Efstathiou, G. *Astrophys. J.* **313**, 505-517 (1987).
- Couchman, H. M. P. & Rees, M. J. *Mon. Not. R. astr. Soc.* **221**, 53-62 (1986).
- Braun, E., Dekel, A. & Shapiro, P. *Astrophys. J.* (submitted).
- White, S. D. M., Davis, M. & Frenk, C. S. *Mon. Not. R. astr. Soc.* **209**, 27P-31P (1984).
- Bond, J. R., Centrella, J., Szalay, A. S. & Wilson, J. R. *Mon. Not. R. astr. Soc.* **210**, 515-545 (1984).
- Dekel, A. *Astrophys. J.* **264**, 373-391 (1983).
- Dekel, A. *Astr. Astrophys.* **101**, 79-87 (1981).
- Wasserman, I. *Astrophys. J.* **248**, 1-12 (1981).
- Valdarnini, R. & Bonometto, S. *Astr. Astrophys.* **146**, 235-241 (1985).
- Dekel, A. in *Proc. Eighth Johns Hopkins Workshop on Current Problems in Particle Theory* (eds Domokos, G. & Kovesi-Domokos, S.) 191-199 (World Publishing, Singapore, 1984).
- Dekel, A. *Astrophys. J.* **284**, 445-456 (1984).
- Ostriker, J. P. & Cowie, L. *Astrophys. J.* **243**, L127-L131 (1981).

48. Ikeuchi, S. *Publ. astr. Soc. Japan* **33**, 211-222 (1981).
49. Saarinen, S., Dekel, A. & Carr, B. *Nature* **325**, 598-601 (1987).
50. Bahcall, N. & Soneira, R. *Astrophys. J.* **270**, 20-38 (1983).
51. Klypin, A. A. & Kopylov, A. I. *Soviet Astr. Lett.* **9**, 41-44 (1983).
52. Bahcall, N. in *Nearly Normal Galaxies* (ed. Faber, S.) (Springer, New York, in the press).
53. Ling, E. N., Frenk, C. S. & Barrow, J. *Mon. Not. R. astr. Soc.* **213**, 21P-27P (1986).
54. Dressler, A. *et al. Astrophys. J.* **313**, L37-L42 (1987).
55. Collins, A., Joseph, R. D. & Robertson, N. A. *Nature* **320**, 506-508 (1986).
56. Barnes, J., Dekel, A., Efstathiou, G. & Frenk, C. S. *Astrophys. J.* **295**, 368-373 (1985).
57. Bardeen, J., Bond, J. R., Kaiser, N. & Szalay, A. *Astrophys. J.*, **304**, 15-61 (1986).
58. Vilenkin, A. *Physics Rep.* **121**, 264-315 (1984).
59. Albrecht, A. & Turok, N. *Phys. Rev. Lett.* **54**, 1868-1871 (1985).
60. Turok, N. in *Nearly Normal Galaxies* (ed. Faber, S.) (Springer, New York, in the press).
61. Primack, J. R., Blumenthal, G. R. & Dekel, A. in *Galaxy Distances and Deviations from Universal Expansion* (eds Madore, B. F. & Tully, R. B.) 265-270 (Reidel, Dordrecht, 1986).
62. Albrecht, A. in *Nearly Normal Galaxies* (ed. Faber, S.) (Springer, New York, in the press).
63. Rees, M. J. in *Dark Matter in the Universe* (eds Kormendy, J. & Knapp, G.) 395-409 (Reidel, Dordrecht, 1986).

ARTICLES

Sensitive measurement of fluctuations in the cosmic microwave background

R. D. Davies*, A. N. Lasenby†, R. A. Watson*, E. J. Daintree*, J. Hopkins*, J. Beckman‡, J. Sanchez-Almeida‡ & R. Rebolo‡

* University of Manchester, Nuffield Radio Astronomy Laboratories, Jodrell Bank, Macclesfield, Cheshire SK11 9DL, UK

† Mullard Radio Astronomy Observatory, Cavendish Laboratory, Madingley Road, Cambridge CB3 0ME, UK

‡ Instituto de Astrofísica de Canarias, 38071 La Oaguna, Tenerife, Spain

Extensive high sensitivity observations of the cosmic microwave background have been made on an angular scale of 8° covering a substantial fraction of the northern sky. An observed anisotropy in the sky emission at a level of $\Delta T/T = 3.7 \times 10^{-5}$ has been detected (T is temperature). This level should strictly be interpreted as an upper limit to the cosmic microwave background fluctuations. It is possibly the direct imprint of density perturbations in the early Universe.

THE cosmic microwave background (CMB) is isotropic to a high degree, reflecting the uniformity of the Universe over all scales at the epoch of recombination (redshift $z \sim 1,000$). Nevertheless, if the structure in the present-day Universe has grown from seed fluctuations in the early Universe, then a significant level of inhomogeneity must exist in the matter (and radiation) distribution at recombination. These fluctuations in the radiation field are predicted by adiabatic models to exist in the $\Delta T/T$ range 10^{-4} to 10^{-5} depending on the angular scale and the value of Ω (the ratio of the actual density of the Universe to that required for closure).

Upper limits to the fluctuation amplitude have been obtained by various sensitive experiments in the range $\Delta T/T = 3 \times 10^{-5}$ to 3×10^{-4} on arc min scales¹⁻³. These limits appear to rule out adiabatic models in which the present mean density of the Universe is composed of normal (baryonic) matter⁴⁻⁶. Low density ($\Omega < 0.1$) models of the Universe dominated by non-baryonic dark matter⁷⁻⁹ are also open to doubt. But, it is possible that fluctuations on arc min scales may be erased in a low redshift ($z = 3-10$) re-ionization phase which can occur during galaxy formation. Fluctuations on angular scales $> 5^\circ$ are not expected to be erased by the re-ionization phase¹⁰.

Accordingly we have chosen to search for fluctuations in the CMB on a scale of $\sim 10^\circ$. This scale is not only unaffected by scattering from the small angle re-ionization structure but has the additional important property that it is larger than the horizon scale at recombination ($\theta \sim 2^\circ$). In the standard Big Bang model causal processes in the early Universe cannot remove anisotropy on these scales.

Evidence for structure approaching a scale of 10° at recombination is becoming available. Superclusters, voids and associated structures^{11,12} are found up to 100 Mpc in extent; this linear size corresponds to $\sim 1^\circ$ at recombination. Larger-scale structures are confidently expected and are even indicated by the distribution of clusters of clusters.

A further reason for choosing a scale of 10° is that a scale-invariant spectrum of fluctuations is predicted by inflationary models; cold dark matter models imply coherence angles for the microwave background fluctuations which are much larger than in previous adiabatic models. The most sensitive tests of these models come from measurements on scales of degrees rather than arc min (G. Efstathiou, personal communication).

The observing strategy

Searches for fluctuations at the sensitivity levels required, 1 mK and below, demand careful experiments. Here, the expected signals are 10^{-6} – 10^{-5} of the receiver system noise. Firstly, low system noise and a high stability are required. Secondly, atmospheric noise arising from water-vapour structure in the lower atmosphere moving through the beam can also produce variations in the receiver output. Our experiment was designed to obviate these problems as far as possible. Low-noise cryogenically-cooled receivers having an overall system noise in the range 80–100 K were used. The receiver performance was continuously monitored with a fast-switching calibration signal. Fast beam-switching between adjacent sky positions also helped overcome receiver stability problems. We chose a high dry site at Izana, Tenerife, at an altitude of 2,300 m to reduce atmospheric problems. A frequency of observation of 10.4 GHz was adopted; it was low enough to be scarcely affected by atmospheric water-vapour fluctuations and high enough to be only a little affected by galactic synchrotron and thermal emission. The effects of 'spillover' surrounding the observing beam, which plague searches for arc min-scale fluctuations using steerable telescopes^{3,13,14,15} were eliminated by using a fixed telescope whose surroundings can be made sanitary from a radio viewpoint. The ultimate improvement to system performance came from using a three-beam arrangement involving the switched double-beam being 'wagged' in azimuth (Fig. 1). This improved the long-term

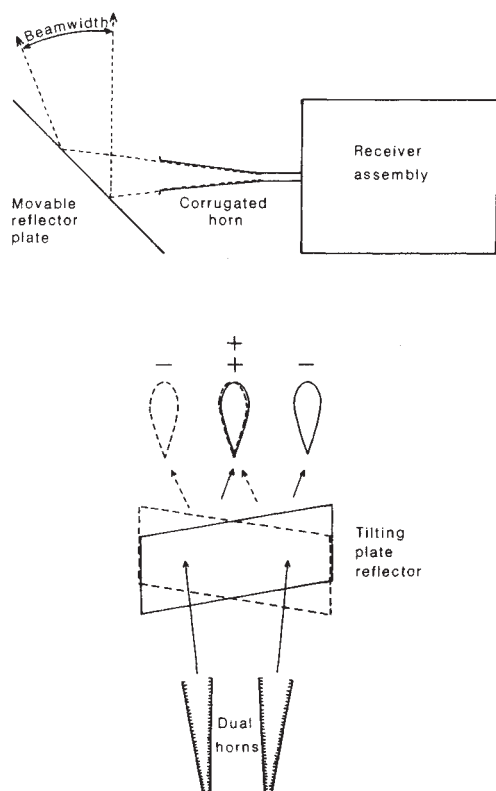


Fig. 1 Arrangement of the triple-beam radiometer for measuring fluctuations in the CMB on a scale of 8° at a frequency of 10.4 GHz. Each of the two horns has a FWHM beamwidth of 8.5° ; their axes are offset by 8.2° . They each receive signals from the sky through a reflector plate which can be moved to allow scans to be made at different declinations. The two beams can be moved in azimuth (hour angle) by tilting the reflector about an axis in its plane. This tilt is arranged so as to swing the beams by 8.2° in the sky. By suitably phasing the switch cycle the total beam response consists of a central positive beam with a negative beam of half the amplitude displaced 8.2° either side.

(several hours) stability by a factor approaching 10. Our final stacked data on beamwidth scales appeared not to be limited in signal-to-noise by longer-term drifts as encountered in most experiments of this type, but rather by short-term noise arising from the receivers and on occasions by an additional contribution from atmospheric irregularities. The fact that separate data streams were recorded by two independent receivers enables us to confirm that long-term drifts were not affecting our results.

The observing programme consisted of two parts. The first was a large-area survey of the northern sky which consisted of 24-h scans in right ascension (RA) at 5° intervals in declination, δ , between $\delta = -15^\circ$ and $+55^\circ$. The second was a deep study of the RA bands at $\delta = 0^\circ$ and $\delta = 40^\circ$ which each consisted of 20 repeated scans in good weather. These scans were stacked to provide about five times more sensitivity than in a single scan. Adjacent scans at δ -offsets of $\pm 5^\circ$ on each side of these scans were also made and were used to add sensitivity to the two central scans.

Observations

Here, we discuss the observations taken at $\delta = 40^\circ$. The data obtained in good atmospheric conditions were stacked to give the added scan in Fig. 2. Only 20% of the observations were rejected because of detectable atmospheric effects. Typically, 20 scans contributed to each point between RA = 09 h and 24 h (135° – 360°). The greater noise between RA = 00 h and 08 h was

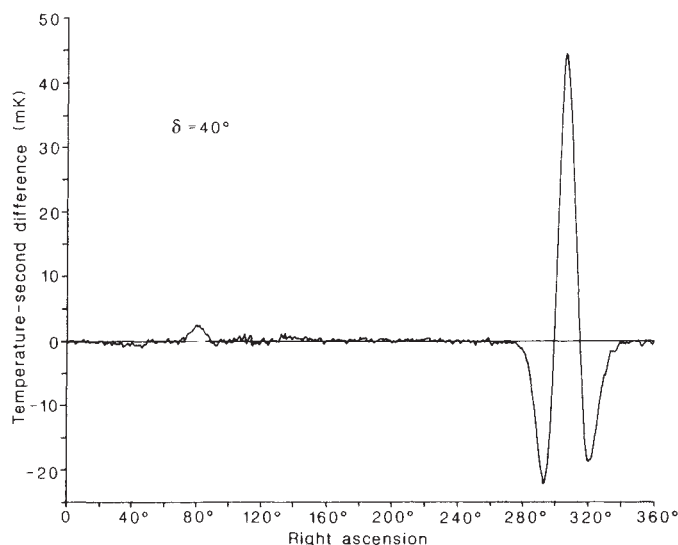


Fig. 2 Averaged output of the beam-switching receiver for the track around the sky at $\delta = 40^\circ$. The triple-beam response can be gauged from the response to the strong galactic plane crossing at RA = 20 h 20 min (305°). The weaker crossing at RA = 03 h 30 min (52.5°) to 05 h 20 min (80°) can also be seen. Data have been averaged over RA intervals of 4 minutes.

the result of fewer scans in this daytime observing period when daily calibrations and reductions were made; weak effects from the Sun may also contribute to this increased noise. Data from the two independent receivers have been added. Figure 2 clearly shows the galactic plane crossings at RA = 04 h and 20 h the second of which includes the extragalactic radio source Cygnus A and the thermal galactic region Cygnus X which together act as a secondary calibrator on each day. For analysis we chose the data between RA = 12 h and 17 h which was observed during the night-time and was also more than 30° from the galactic plane. The scan of this region, smoothed over 4-min intervals in RA, is shown on an expanded scale in Fig. 3. A linear baseline has been removed from these data; this has the effect of reducing our sensitivity on scales of 35° or more. The beam response is shown for comparison.

Analysis and assessment of the results

In our discussion of the results of this experiment, we need to assess how much of the structure in Fig. 3 is attributable to the CMB and how much to receiver and atmospheric noise. Our approach follows that of Lasenby and Davies³, in that day-to-day variations are used to assess the measurement error, but the detailed separation is made using a maximum likelihood calculation which takes into account the geometry of the beams. Details of this calculation will be published separately. In this approach we model the measured temperature differences as a (column) vector $\mathbf{y} = (y_1, y_2, \dots, y_n)^T$ (with $n = 70$ for the part of the $\delta = 40^\circ$ data used) equal to the sum of a true signal $\mathbf{x}$ and a gaussian error contribution $\mathbf{\epsilon}$. The true signal $\mathbf{x}$ is in turn taken as a random gaussian vector with covariance matrix $\mathbf{M}$ determined by the true sky autocovariance function (a.c.f.) $C(\theta)$ and the size and geometry of the beams. ($C(\theta)$ is defined as the expected value of $\Delta T(\mathbf{n})\Delta T(\mathbf{n}')$ for directions $\mathbf{n}$ and $\mathbf{n}'$ satisfying $\mathbf{n} \cdot \mathbf{n}' = \cos \theta$.) A simple gaussian form was taken for $C(\theta)$ equal to $C_0 \exp(-\theta^2/2\theta_c^2)$ where C_0 is the true sky variance (as would be measured with a δ -function beam) and θ_c is a sky coherence scale. Notice that when folded with a single gaussian beam of dispersion σ ($= 3.5^\circ$ in our case) the true sky a.c.f. becomes modified to $C_M(\theta) = C_{M_0} \exp(-\theta^2/2(2\sigma^2 + \theta_c^2))$, where $C_{M_0} =$

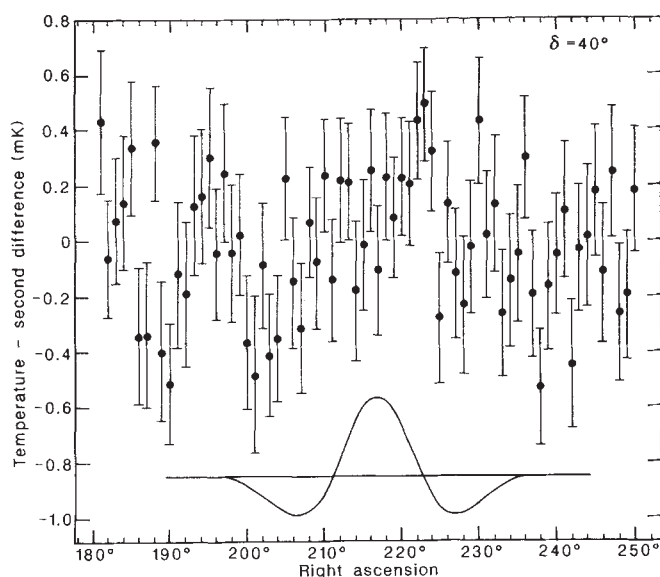


Fig. 3 The averaged output from the beamswitch receiver for the RA range 12 h to 17 h at $\delta = 40^\circ$. This section of the track was observed at night and lay more than 30° from the galactic plane. Data are binned into 4-min intervals of RA. The triple-beam response is illustrated.

$C_0\theta_c^2/(2\sigma^2 + \theta_c^2)$ is the diluted variance as seen in a beam of this size.

A maximum likelihood calculation is then performed on the observed data with the form of the true sky a.c.f. $C(\theta)$ given above and the covariance matrix M describing the beam geometry. The parameters C_0 and θ_c characterizing $C(\theta)$ are then varied, enabling likelihood contours to be constructed in the C_0, θ_c plane. This analysis showed that fluctuations of astronomical origin were detected—a non-zero variance was preferred over zero variance by a factor of approximately 10 to 1 in likelihood. A range of angular scales was selected consistent with the experimental geometry, and we discuss here the results for the angular scale on which the experiment had greatest sensitivity, which was revealed by the maximum likelihood analysis to be 8° – 10° FWHM (full width at half maximum), corresponding to $\theta_c \approx 4^\circ$. Assuming this scale for the sky coherence function we find that within our 8.3° beam we are seeing an excess standard deviation of $\sqrt{C_{M_0}} = 0.10$ mK which is attributable to emission from the sky. (Taking $\theta_c \neq 4^\circ$ gives larger values for the maximum likelihood value of $\sqrt{C_{M_0}}$.) This value can be compared with an upper limit of 1.5 mK given by Mandolesi *et al.*¹⁶ (but see below for a discussion of their results). Our result corresponds to a positive detection of fluctuations at a level of $\Delta T/T = 3.7 \times 10^{-5}$. Note that this result is not particularly sensitive to the accuracy of the derived values of the instrumental errors σ_i . We find that these errors have to be multiplied by the large factor of 1.8 before the likelihood contours become peaked at zero rather than at positive values of C_0 . This behaviour reflects the fact that the likelihood function automatically includes all the correlation information available, and that, in the presence of a real signal, this forms a powerful constraint, not easily overwhelmed by increases in the assumed errors (N. Kaiser and A.N.L., in preparation). This is quite unlike the behaviour of other estimators of sky variance which have been used in the past, which have relied wholly on a comparison of internal against external variance to estimate the contribution from the sky (see Appendix B.2 in Lasenby and Davies³).

The intrinsic fluctuation amplitude $\sqrt{C_0}$ corresponding to the value of $\sqrt{C_{M_0}} = 0.10$ mK found above, is 0.16 mK, or 5.7×10^{-5}

in $\Delta T/T$, again assuming $\theta_c = 4^\circ$. It should be remembered that observers generally do not attempt to deconvolve the effects of their beam in this way, and instead quote a value of (or an upper limit on) what is in effect the diluted r.m.s. $\sqrt{C_{M_0}}$, or some combination of $C_M(\theta)$ s involving the beam-switching geometry, either explicitly or implicitly. This is discussed further in the next section.

We should point out that our detection of sky fluctuations may contain a component of high-latitude radio continuum emission from our Galaxy. Surveys of the Galaxy at 0.408¹⁷ and 1.4 GHz¹⁸ show a low level of structure at high latitudes which could make a significant contribution to the observed sky fluctuations, depending on the spectral index of the synchrotron radiation. A preliminary investigation in collaboration with Dr J. L. Osborne indicates that the observed structure at 1.4 GHz would lead to an r.m.s. fluctuation of 0.07 mK at 10.4 GHz for a temperature spectral index of $\alpha = 2.8$ ($T(\lambda) \propto \lambda^{-\alpha}$). However, there is no match between the shape of the observed emission and the structure predicted from the 1.4-GHz data; neither is there a match with the 0.408-GHz data. In any case, a spectral index of 3.0 or greater may be appropriate at these frequencies; an index of 3.0 would lead to a reduced galactic contribution of 0.05 mK. Further, a spectral index range greater than that observed between 0.408 and 1.4 GHz would be required to explain the difference between the structure at low frequencies and 10.4 GHz. High galactic latitude H α emission of the type measured by Reynolds¹⁹ would contribute a mean temperature of 0.02 mK; the r.m.s. value smoothed over our 8° beam would be significantly less than this. In view of these considerations we believe that a substantial part of the detected signal may be due to CMB anisotropy, given reasonable assumptions about the spectrum of the galactic background.

Discussion

We have described the highest sensitivity experiment yet performed in the search for CMB fluctuations on scales of 5° – 15° , reaching an equipmental sensitivity (standard deviation) of 0.07 mK, using repeated observations along RA tracks. Significant fluctuations with an observed standard deviation of 3.7×10^{-5} were found. This amplitude corresponds to intrinsic sky fluctuations of $\Delta T/T = 5.7 \times 10^{-5}$ on an angular scale of $\sim 8^\circ$ – 10° . If there is a galactic contribution, then these values are upper limits on CMB fluctuations.

These results bear directly on several important aspects of the large-scale structure of the Universe. A detailed discussion, together with a fuller comparison with other experimental work, will be published separately, but the main conclusions are presented here.

In common with other recent measurements of microwave background anisotropy (such as ref. 1) the current results rule out models of galaxy formation via adiabatic fluctuations in which there is no 'dark matter' component (that is, standard baryonic models). This applies for all reasonable values of Ω and H (Hubble's constant). Using the predictions given in refs 6 and 20 and adopting their a.c.f., we find that the only universes which survive are those with high density ($\Omega = 1$) and high-power-law slope ($n > 1$). The strongest constraint is placed by the Uson and Wilkinson result for arc min scales. But, if reionization occurs at a late epoch ($z = 5$ – 10) this arc min-scale result is nullified and the present result sets the constraint. In this case lower density ($\Omega \leq 0.3$) scale-invariant ($n = 1$) models are still eliminated. These include the canonical (low-density) baryonic model.

Partly as a result of these microwave background experiments, the 'standard picture' has now become one dominated by cold dark matter (CDM), rather than baryons (thus reducing the small-scale anisotropy), and with the further addition of 'inflation' to provide a causal explanation for isotropy on scales above the horizon scale at decoupling. Following naturally from inflation is the prediction that the initial fluctuations have a scale-

invariant form and that $\Omega = 1$. Accepting this standard picture leads to fairly unambiguous predictions for the level of microwave anisotropy expected, against which our present results may be compared. Such a comparison also allows the relative sensitivity of our experiment to be judged, once factors such as beam size and switching geometry have also been taken into account. Accordingly, we have calculated the $\Delta T/T$, expressed as an r.m.s., which would be observed with our beam configuration from a standard CDM model. The model used is that described in G. Efstathiou and J. R. Bond (personal communication) and has the parameters $\Omega = 1$, $H = 75 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $\Omega_{\text{baryon}} = 0.03$. The amplitude of the perturbation spectrum is fixed by comparison with the second moment of the empirically observed galaxy correlation function, J_2 . For the resulting anisotropy covariance function $C(\theta)$ we have used a two component analytic fit (kindly provided by Dr G. Efstathiou) which should preserve the essential features of the original. This function has been folded with the beam as described above to provide a single beam autocovariance function $C_M(\theta)$, and then combinations of the $C_M(\theta)$ s taken as appropriate to the switching geometry are made to find a predicted measured r.m.s. The r.m.s. fluctuation predicted from this model is a factor of 8 below our observed result. This 'gap' measures the degree to which our experiment puts pressure on the standard CDM model. Of course if our measured r.m.s. contains a galactic contribution then the intrinsic CMB fluctuations will be lower and the corresponding pressure on the CDM model will be less. Carrying out a similar analysis for the Uson and Wilkinson^{1,2} experiment which achieved a similar noise level but for which the result was an upper limit rather than a detection we find that the 'gap' is a factor of 13 rather than 8. This is compatible with our detection and reflects the large coherence scale for CDM fluctuations.

It is worth making some comment on the three published results on the scale of a few degrees. (1) If the Mandolesi *et al.*¹⁶ data are examined in more detail (N. Kaiser and A.N.L., in preparation), it is found that a signal compatible with their beam geometry is evident, which appears to correspond to a detection of observed fluctuations with r.m.s. $\Delta T/T = 4 \times 10^{-4}$, which is approximately 10 times our detection level. We believe that this figure is not representative of the general level of sky fluctuations as such high amplitudes are not seen in the large area covered by our survey. Moreover, the disagreement with the limits imposed by other experiments is obviously acute and would require the fluctuation spectrum to peak sharply on a scale of 2° , which is unlikely on current models. (2) The results of Melchiorri *et al.*^{21,22} need careful interpretation because few details of either their experimental configuration (in particular their beam patterns) or their statistical procedures are available. Further, their result has (like ours) to be taken as either an upper limit or a detection depending on an uncertain galactic contribution (in this case from high latitude dust emission as exemplified by Infrared Astronomical Satellite cirrus). (3) The result of Fixen *et al.*²³, which constrains $C_M(\theta)$ to be less than 0.01 mK^2 for scales $10^\circ \leq \theta \leq 180^\circ$, is of a different character

from the others since it refers directly to the value of the covariance function at non-zero lag. No confidence level is given for this result and no details are given of how the sky map (from which this limit is derived) was reconstructed from the 90° difference data. This makes it difficult to compare with other results.

On scales $\geq 2^\circ$, the predictions for $\Delta T/T$ are dominated by the Sachs-Wolfe effect from perturbations which are yet to become nonlinear. As already mentioned, it is highly unlikely that they can be erased by subsequent re-ionization. Hence a measurement of $\Delta T/T$ on these scales can tell us directly about the density distribution at recombination in a way impossible for the smaller-scale results (for instance, those of Uson and Wilkinson refs 1, 2). Assuming a power-law spectrum for the initial fluctuations $k(|\delta_k|^2 \propto k^n)$ of amplitude δ_k at a wavenumber means that given a measurement of $\Delta T/T$ on a single angular scale $\geq 5^\circ$ one can predict the values which should be observed on all other scales up to quadrupole and dipole. This point is noted and used for example by Peebles¹⁰ and by Abbott and Wise²⁴, who both assume a scale-invariant ($n=1$) spectrum, with the latter authors relating these amplitudes to a single dimensionless number ϵ_H (the relative amplitude of density perturbations on re-entering the horizon scale). Adopting a similar viewpoint, and working from our measured value at $\sim 8^\circ$ to derive the normalization on an $n=1$ spectrum, we find that the quadrupole amplitude a_2 is predicted to be $\leq 4.3 \times 10^{-5}$, whereas $\epsilon_H \leq 1.9 \times 10^{-5}$. In both cases the upper bounds correspond to assuming that our detection is entirely due to the CMB. The predicted quadrupole amplitude may be compared with a recent upper limit (ref. 25; G. Efstathiou and J. R. Bond, personal communication) on a_2 of $< 1.1 \times 10^{-4}$.

If Ω is not given by the inflationary value $\Omega = 1$, but satisfies $\Omega < 1$, then the situation is more complex. A new effect enters on intermediate angular scales due to focusing of the CMB into one or more hotspots of characteristic dimension Ω radians (for example, ref. 26). Here limits on the possible existence of such hotspots depend on the fraction of sky searched. With the extension of our analysis to overlapping declination strips from -15° to $+45^\circ$, it will be possible for us to put strong constraints on such spots (and hence on Ω), as also on the distribution of strings (for example ref. 27), whose signatures must again be searched for over a wide area.

Finally we note the work of Goicoechea and Sanz²⁸, on the existence of isolated spherical voids in the Universe, and the limits that intermediate-scale anisotropy place upon these. Interpreting our result as a limit, we are able to reduce their upper bound on possible isolated mass fluctuations $|\delta M/M|$ from $\leq 10\%$ to $\leq 1\%$, thus strengthening their conclusion that large-scale structures cannot be isolated.

In view of the strength of our results in confining models of the evolution of the Universe, we are continuing to make sensitive measurements on a 5° scale; observations are also being made at 5 GHz to delineate more quantitatively any galactic contribution to the observed sky fluctuations.

Received 19 January; accepted 27 February 1987.

- Uson, J. & Wilkinson, D. T. *Nature* **312**, 427-429 (1984).
- Uson, J. & Wilkinson, D. T. *Astrophys. J.* **283**, 471-478 (1984).
- Lasenby, A. N. & Davies, R. D. *Mon. Not. R. astr. Soc.* **203**, 1137-1169 (1983).
- Silk, J. *Astrophys. J.* **151**, 459-471 (1986).
- Peebles, P. J. E. & Yu, J. T. *Astrophys. J.* **162**, 815-836 (1970).
- Wilson, M. L. & Silk, J. *Astrophys. J.* **243**, 14-25 (1981).
- Peebles, P. J. E. *Astrophys. J.* **263**, L1-L5 (1982).
- Bond, J. R. & Efstathiou, G. *Astrophys. J.* **285**, L45-L48 (1984).
- Vittorio, N. & Silk, J. *Astrophys. J.* **285**, L39-L43 (1984).
- Peebles, P. J. E. *Astrophys. J.* **243**, L119-L122 (1981).
- De Lapparent, V., Geller, M. & Huchra, J. *Astrophys. J.* **302**, L1-L5 (1985).
- Oort, J. H. A. *Rev. Astr. Astrophys.* **21**, 373-428 (1983).
- Lake, G. & Partridge, R. B. *Astrophys. J.* **237**, 378-389 (1980).
- Birkinshaw, M. & Gull, S. F. *Mon. Not. R. astr. Soc.* **206**, 359-375 (1984).
- Chase, S. T., Joseph, R. D., Robertson, N. A. & Ade, P. A. R. *Mon. Not. R. astr. Soc.* (in the press).

- Mandolesi, N. *et al. Nature* **319**, 751-753 (1986).
- Haslam, C. G. T., Salter, C. J., Stoffel, H. & Wilson, W. E. *Astr. Astrophys. Suppl.* **47**, 1-140 (1982).
- Reich, P. & Reich, W. *Astr. Astrophys. Suppl.* **63**, 205-292 (1986).
- Reynolds, R. J. *Astrophys. J.* **282**, 191-196 (1984).
- Wilson, M. L. *Astrophys. J.* **273**, 2-15 (1982).
- Melchiorri, F., Melchiorri, B. O., Ceccarelli, C. & Pietranera, L. *Astrophys. J.* **250**, L1-L4 (1981).
- Fabbri, R., Guidi, I., Melchiorri, F. & Natale, V. *Phys. Rev. Lett.* **44**, 1563-1566 (1980).
- Fixen, D. J., Cheng, E. S. & Wilkinson, D. T. *Phys. Rev. Lett.* **50**, 620-622 (1983).
- Abbott, L. F. & Wise, M. B. *Astrophys. J.* **282**, L47-L50 (1984).
- Lubin, P., Villela, T., Epstein, G. & Smoot, G. *Astrophys. J.* **298**, L1-L5 (1985).
- Bajtlik, S., Juszkiewicz, R., Proszynski, M. & Amsterdamski, P. *Astrophys. J.* **300**, 463-473 (1986).
- Kaiser, N. & Stebbins, A. *Nature* **310**, 391-393 (1984).
- Goicoechea, L. J. & Sanz, J. L. *Astrophys. J.* **293**, 17-24 (1985).

Differentiated parental DNA strands confer developmental asymmetry on daughter cells in fission yeast

Amar J. S. Klar

Cold Spring Harbor Laboratory, PO Box 100, Cold Spring Harbor, New York 11724, USA

The two strands of the DNA molecule are complementary but not identical. Hence, upon semiconservative replication, different parental DNA strands are segregated to daughter cells. A molecular analysis suggests that the process of fission yeast mating-type interconversion uses asymmetry of the DNA strands to generate a regular lineage of cellular differentiation.

How different cell types are generated in a sequential and programmed fashion in a developing organism remains a major unanswered question. The signal for cell type determination may be asymmetrically distributed to daughter cells through the cytoplasm¹ or the genome². I address this issue by studying the case of fission yeast, in which two daughter cells regularly acquire different developmental potential.

As a model system, I investigated cell type changes in the fission yeast *Schizosaccharomyces pombe*. This yeast generally exists in a haploid state and has two mating-cell types, called P (for Plus) and M (for Minus), controlled by alternate alleles of the *mat1* locus³. These alleles efficiently interchange (switch) and follow a remarkable, nonrandom pattern of switching in a cell lineage. Miyata and Miyata⁴ have demonstrated a developmental asymmetry; of a pair of sister cells, only one member produces a single switched cell in the next generation in at least 74% (with a maximum of 92%) of cell divisions (Fig. 1). In the remaining 8 to 26% cell divisions, no switches are found, but their progeny can switch in subsequent generations. In other words, only one grandchild is switched of the four progeny produced by two cell divisions of a single cell ('Miyata's rule'; Fig. 1). The switches are shown to occur by a transposition-substitution event in which a replica of either the *mat2-P* (containing Plus information) or *mat3-M* (containing Minus information) donor 'cassette' is transmitted to the *mat1* locus via a gene conversion process. The cassette at *mat1* is transcriptionally active and determines the cell type, whereas those at *mat2* and *mat3* are transcriptionally inert and act only as donors of genetic information⁵⁻⁸ (Fig. 2).

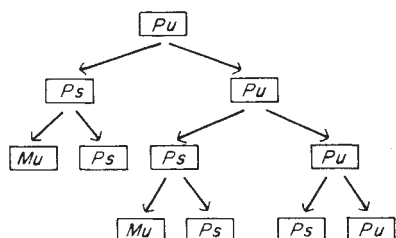


Fig. 1 The Miyata's⁴ pattern of *mat1* switching in a cell lineage showing the switching rule of one-in-four related cells. The *Mu* and *Pu* (*u* for unswitchable) designations indicate recently switched cells that cannot produce switched cells in the next generation, while the *Ms* and *Ps* (*s* for switchable) cells can. The sister of the recently switched cell can efficiently produce a single switched daughter in the next generation (R. Egel, personal communication). This feature suggests that the recently switched cell must not switch in the next generation, because such a switch would violate the one-in-four rule. See text for details.

In this article I address the basis of this developmental asymmetry: why does only one member of a pair of sister cells produce switched progeny? Specifically, I examine whether the ability to switch is inherited through the chromosome or due to cytoplasmic (plus nuclear) factors unequivocally transferred to daughter cells. In a study of switching of diploid *S. pombe* cells, Egel⁹ demonstrated that the asymmetrically segregating competence for switching is separately determined for the two homologues, indicating a chromosomal basis for the segregation of developmental asymmetry. By an independent procedure, I also show that the competence for switching is acquired by an 'epigenetic' chromosomal marking, known as 'imprinting'¹⁰. I show that the site of imprinting is the *mat1* locus, and that the developmental asymmetry is a consequence of only one of the two chains of the DNA molecule (which are complementary, but not identical¹¹) acquiring the developmental potential.

Switch-defective mutants

Previous studies have shown that the recombination event required for switching is initiated by an *in vivo*-generated double-stranded DNA break at *mat1*^{6,12}. In a steady-state culture, ~20% of the cells have a break at the *smt* (switch of mating type) site (Fig. 2). The observed level of the break is constant throughout the cell cycle. The switching event is a mitotic gene conversion event that repairs (heals) the break. Strains deleted for the *mat2* and *mat3* donor loci have the break, maintain a stable mating-type, and do not have dead progeny¹³. Therefore, this yeast is capable of healing the double-stranded break at *mat1* without switching. Several switching-defective (*swi*⁻)

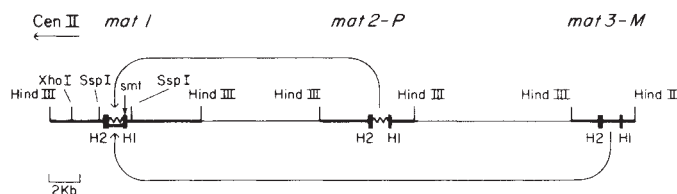


Fig. 2 A portion of chromosome II showing the mating-type region. Thick lines show the 10.4-kb 6.3-kb and 4.2-kb *Hind*III fragments containing *mat1*, *mat2-P* and *mat3-M*, respectively. The regions of homology common to each cassette (H1, 59-bp) and (H2, 135-bp) are represented by blocks bordering the *P*-specific 1,113-bp (~) or *M*-specific 1,127-bp (-) regions (M. Smith, personal communication). The designation *smt* (switch of mating-type signal) marks the site of the double-stranded DNA cut. The arrow (Cen II) indicates the direction of the centromere. The *mat2* gene is situated about 17-kb away from *mat1*, and *mat3* is located about 15-kb distal to *mat2*⁷. A unique *Bam*HI site is present in the *P* cassette and a unique *Eco*RI site is present in the *M* cassette.

Fig. 3 Strand segregation model. Details of the model are presented in the text. *P* and *M* indicate the *mat1* alleles; *W* and *C* denote the Watson and Crick strands of DNA; * indicates the suggested strand-specific imprinted event and ~ the strand which can be imprinted. Arrows originating from the grandparent cell indicate segregation of a particular DNA chain to the daughter cell. The sister of the recently switched cell (second grandchild from the left) is supposed to inherit the cleaved chromosome as it is known to be able to switch in the next generation (R. Egel, personal communication). A gap in the continuity of the chromosome indicates the presence of a double-stranded DNA break. *P_u*, *P_s*, *M_u* and *M_s* are defined in the legend to Fig. 1.

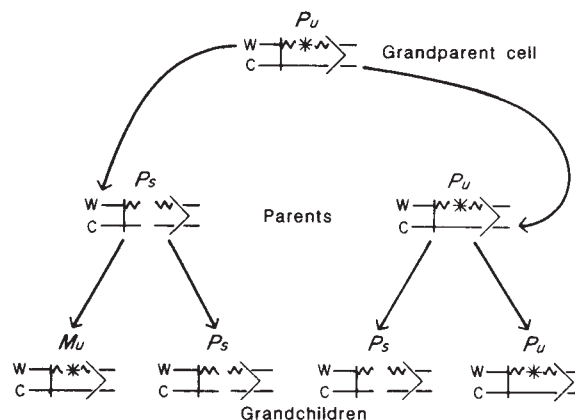


Table 1 Experimental strains of *S. pombe*

Strain	Genotype
SP112	<i>mat1-M or-P mat2-P mat3-M (h⁹⁰) swi7 leu1-32 ade6-M216</i>
SP401	<i>mat1-M or-P mat2-P mat3-M (h⁹⁰) swi5 leu1-32 his3</i>
SP487	<i>mat1 direct duplication::LEU2 mat2-P mat3-M swi5 leu1-32</i>
SP488	<i>mat1 inverted duplication::LEU2 mat2-P mat3-M swi5 his3 leu1-32</i>

mutants were characterized: the *swi1*, *swi3* and *swi7* gene functions were found to be required for the formation of the break *swi2*, *swi5* and *swi6* to be needed to use the break for recombination and *swi4*, *swi8*, *swi9* and *swi10* functions for resolution of the switching intermediates^{12,13}.

DNA strand segregation model

Why does only one of four descendants of a given chromosome switch? The decision for a given switch must have been made two generations earlier by the 'grandparent' cell (Fig. 1). We may imagine a molecular model in which the 'Watson and Crick' strands of DNA¹¹ are not equivalent in their ability to acquire the developmental potential for switching. Specifically, we propose that some *swi* gene functions catalyse an imprinting event, possibly a sequence-specific DNA modification, such that only one specific chain (say, the Watson strand) at the *mat1* locus is modified (Fig. 3). We imagine that the site of modification is not palindromic. Thus one of the two strands can be modified specifically, in a fashion analogous to that catalysed by the *HinfIII* type III restriction-modification enzyme¹⁴. It seems likely that the DNA modification is required for generating a double-stranded DNA break in the chromosome of its progeny. Then replication of the hemi-modified chromosome generates two sister chromatids, one of which is cut (the one possessing the old Watson chain) and the other uncut. The cell inheriting the broken chromosome would produce a single switched cell in the next generation. Although the specific details of the imprinting event are not known, the key idea, of the segregation of DNA strands conferring developmental asymmetry to daughter cells, is testable.

Similar theoretical mechanisms of strand-specific modification and segregation have previously been proposed as an explanation for cellular differentiation², for perpetuating stem cells while protecting them against mutation¹⁵, as well as for the yeast mating-type system⁹.

Construction of *mat1* duplications

A molecular test of this strand segregation model consists of determining the effect of cassette orientation on switching of tandem cassette duplications of *mat1*. Two constructions, one containing a directly oriented and the other an inversely oriented

cassette duplication at *mat1*, were made (see Fig. 4 legend) and placed in the genome. Both constructions were cut out of the vector at the terminal *HindIII* and *BamHI* sites. The linear DNA was used for DNA-mediated transformation¹⁶ of strain SP112 (*h⁹⁰ swi7⁻ leu1⁻*), and *Leu⁺* transformants were selected. (The *Saccharomyces cerevisiae* *LEU2* gene present in each construction is known to complement the *Saccharomyces cerevisiae* *leu1⁻ defect¹⁶*.) The switching-deficient *swi7⁻* strain was used as the recipient because we could not predict the stability of the *mat1* duplications in an efficiently switching wild-type strain. In another study a deletion covering the *XhoI* site, into which were inserted the *mat1* and the *LEU2* sequences, was found to be haploid-viable (R. Cafferkey and A.J.S. K., unpublished data). Therefore, transformation of only the haploid strain was attempted to place the duplication in the genome. Homologous recombination between the ends of the 8.3-kilobase (kb) *HindIII*-*BamHI* fragment with the homologous sequences in the chromosome¹⁷ would insert *mat1* as well as *LEU2* sequences in the genome at the *mat1* locus to generate the desired *mat1* duplications. A similar procedure was employed to delete *mat2* and *mat3* (ref. 13; see also refs 18 and 19). The desired genomic constructions should increase the size of the *mat1*-containing *HindIII* fragment from 10.4 kb to 14.1 kb. One of 15 transformants generated using directly oriented construction and two of 20 with inversely oriented construction had the desired 14.1-kb *HindIII* fragment and lacked the original 10.4-kb fragment. One example of each of the required genotypes is given in Fig. 4b (see also Table 1). The rest of the transformants also resulted from homologous recombination but contained multiple insertions (data not shown) and were therefore not used. The haploid strains with duplication sporulated azygotically, indicating the expression of both *mat1-P* and *mat1-M* information. Such a phenotype is similar to that of a *pat1* (or *ran1*) mutant that can sporulate in the haploid state^{20,21}.

A test of the model

To study the switching ability of cells harbouring duplicated cassettes, strains with the *swi⁺* genotype were constructed by genetic crosses. The inverted duplication was completely stable in wild-type (*swi⁺*) genotype. The direct duplication was, however, found to be mitotically unstable, and *Leu⁻* segregants that had lost the *mat1* duplication were generated. Nearly 50% of the subclones of the *Leu⁺* colony were found to be *Leu⁻*. Presumably, the instability is due either to substitution of one cassette for both cassettes in one switching event or to the loss of the DNA sequence present between the two *smt* sites (the cut sites; see below). Similar switching of *MAT* duplications to generate a single cassette has been observed in *S. cerevisiae*²². Genetic crosses were carried out to introduce both duplications into a *swi5⁻* background with the hope that in this genotype the *mat1* direct duplication would be more stable. The *swi5⁻* strains are known to have the break but are defective in using

Fig. 4 *a*, An *in vitro* construction used to produce genomic *mat1* duplications. *b*, Southern blot analysis of SP112 transformants containing *mat1* duplications. Lane 1, SP112 (*h⁹⁰swi7⁻*) DNA; lane 2, SP112 transformant DNA containing direct *mat1* duplication; lane 3, SP112 transformant DNA containing inverted *mat1* duplication. The faint 5.4- and 5.0-kb bands result from the *in vivo* cleavage at the *smt* site. This level of break is expected because the *swi7⁻* allele used is leaky³⁴. Numbers indicate sizes of DNA fragments in kilobase pairs (kb). **Methods.** The 8.3-kb fragment was constructed by cloning the 2.2-kb *SalI*-*XhoI* fragment (that contains the *S. cerevisiae* *LEU2* gene) into the *mat1* proximal *XhoI* site contained within the centromere proximal *HindIII*-*BamHI* 4.55-kb fragment, the latter derived from the *mat1-P* cassette containing DNA (Fig. 2). Into the resulting unique *XhoI* site a 1,450-bp *SspI* fragment containing the entire *mat1-P* cassette was cloned after adding *XhoI* linkers to the *SspI* cut blunt ends (Fig. 2). Both orientations of the *mat1* cassette were cloned. These constructions were cloned into the *HindIII*, *BamHI* sites of pBR322. Arrows indicate the orientation of a particular cassette. Standard procedures for manipulating DNA *in vitro* were used³². Methods for culture and genetic crosses have been described³. DNA was isolated from 10-ml saturated cultures as described⁷. For Southern blotting, ~1.0 µg of DNA was digested with the *HindIII* endonuclease and then subjected to electrophoresis through a 0.8% agarose gel. Transfer of DNA from the gel to nitrocellulose paper was as described³³. The probes were nick-translated in the presence of [α -³²P]-labelled deoxyribonucleotides and then hybridized¹⁶. The probe consisted of a 10.4-kb *HindIII* fragment possessing the *mat1-M* cassette cloned into the *HindIII* site of pBR322.

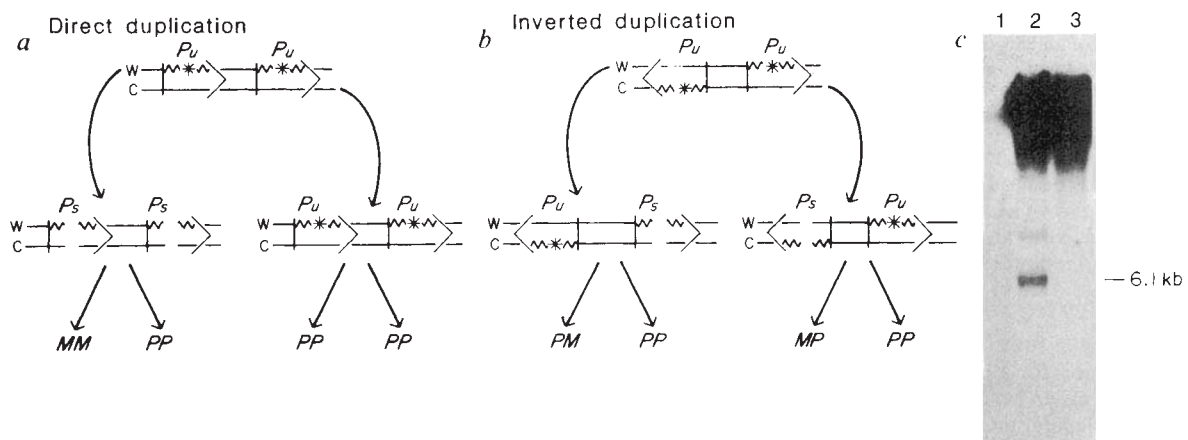
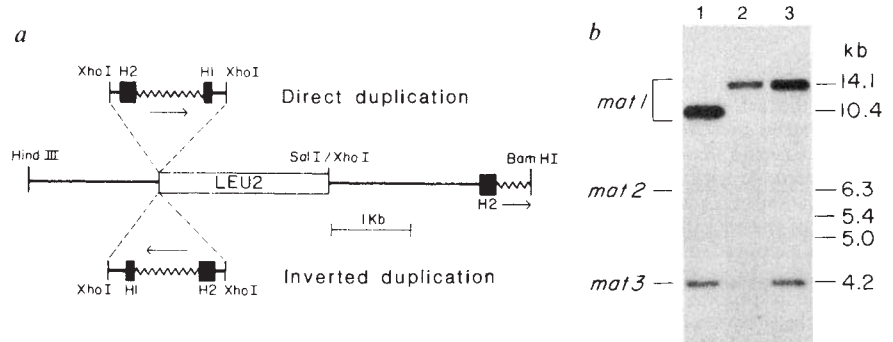


Fig. 5 *a, b*, Predictions of the strand segregation model. See text for details of the predictions. Designations are given in Fig. 3 legend. Wide arrows, cassette orientation. In both cases, the newly placed cassette on the left generated the required duplication. *c*, Southern blot analysis of DNA isolated from strains with direct and inverted *mat1* duplications. Lane 1, SP401 DNA (standard *mat1* allele); lane 2, SP487 DNA (direct duplication); lane 3, SP488 DNA (inverted duplication). Methods are as described (Fig. 4, legend) except that the SP487 and SP488 cells were grown in *Leu⁻* media, the probe consisted of *S. cerevisiae* *LEU2* gene cloned in pBR322, and ~2.5 µg of DNA was loaded in each lane. The *LEU2* sequence has some homology with the *S. pombe* genome, as shown by the level of hybridization to the SP401 DNA (lane 1).

it for recombination and consequently are defective in switching^{12,13}. These mutants are known to be able to heal the break without switching¹³. As will be discussed below, one key test of the model I propose consists only of determining which cassette acquires the double-stranded cut. The *swi5* mutation thus should not affect the analysis.

Both the strand segregation and the cytoplasmic segregation models predict that in the direct tandem duplication a single cell could acquire the competence to switch both cassettes in the same cell cycle (Fig. 5*a*). As nearly 20% of *mat1* DNA is cut *in vivo* and the break persists throughout the length of the cell cycle⁶, we can molecularly monitor the competence of a given cassette to switch by directly assaying for the presence of the double-stranded break at that cassette by the Southern blotting procedure²³. More specifically, the chromosome of the cell designated as *Ps Ps* in Fig. 5*a* should have a double-stranded cut at both *smt* sites to generate a 6.1-kb fragment (the *smt* sites are placed 6.1 kb apart). It should be noted that this prediction will hold true only if the site of imprinting is contained within the duplicated segment of DNA. The predicted fragment was found when the unrestricted DNA isolated from strain SP487

(*mat1* direct duplication) was subjected to Southern blotting and probed with the *LEU2* sequences (Fig. 5*c*, track 2). (The *LEU2* is placed between the *mat1* cassettes, see Fig. 4*a*.) In addition, a larger faint 7.7-kb fragment was also observed. Its origin is unknown but is presently under study. Both of these fragments also hybridized with a probe containing the *mat1* sequence and are also present in *swi5⁺* strains containing the direct duplication (data not shown). This result shows that a haploid cell can acquire the competence to cleave more than one cassette in the same cell cycle. The observed intensity of the 6.1-kb fragment is lower than predicted if both *smt* sites were cleaved at a wild-type rate, possibly because the *smt* site of the newly inserted *mat1* cassette is cleaved at a much reduced rate (see below). Also, the 6.1-kb fragment may be preferentially lost during mitotic growth.

In addition to this physical prediction, it is predicted that a cell with the *Ps Ps* designation in Fig. 5*a* could independently and simultaneously switch both cassettes to generate a *MM* progeny in a single cell cycle. Testing this biological prediction by pedigree analysis is not yet possible because the determination of mating-type of a cell requires its successful mating to

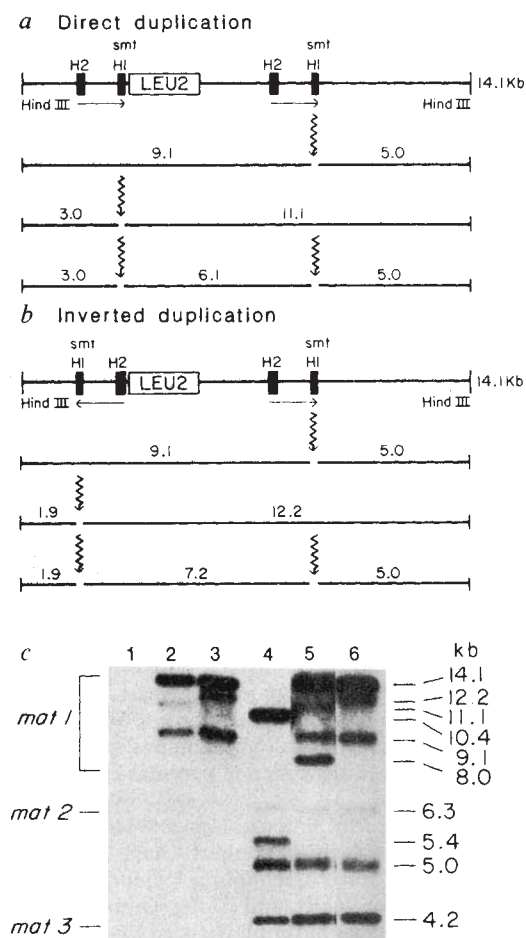


Fig. 6 All possible sizes of the restriction fragments that may be generated by *smi* cleavage in *Hind*III-digested DNA of *a*, direct and *b*, inverted *mat1* duplication. Vertical wavy arrows indicate double-stranded break at a given *smi* site. The numbers indicate sizes of the expected fragments in kb. *c*, Efficiency of DNA cleavage at each *smi* site as assayed by Southern analysis. Lanes 1 and 4, SP401 DNA (standard *mat1* allele); lanes 2 and 5, SP487 DNA (direct duplication); lane 3 and 6, SP488 DNA (inverted duplication). The DNA probe used for lanes 1-3 was pBR322 containing *LEU2*, for lanes 4-6 the 10.4-kb *Hind*III fragment with the *mat1-M* cassette cloned into the *Hind*III site of pBR322 was used. *Hind*III-digested DNA (~1.0 µg) was loaded in each lane. The 1.9-kb and 3.0-kb fragments were run off the gel and are therefore not seen in the figure.

the cell of opposite type, generating a diploid zygote. Consequently, we cannot assay the mating-type of a cell and its mitotic progeny in consecutive generations.

The result presented above is consistent with both models but does show that more than one cassette can become competent to switch in a single cell and within a single cell division cycle. A specific physical prediction that distinguishes between these models is provided by the inverted duplication. According to the cytoplasmic segregation model, the analogous 7.2-kb fragment should be generated in the inversely oriented duplication (the *smi* sites in this construction are 7.2 kb apart), but it should not be generated if the strand segregation model (Fig. 3) is correct. As predicted in Fig. 5b by the strand segregation model, both strands of DNA should be modified, one on one cassette and the other on the second cassette. Segregation of these strands should generate two daughter cells, one having a break in one cassette and the sister inheriting a break in the second cassette. Thus by the strand segregation model it is predicted that a given

cell should switch one or the other but never both cassettes in a single cycle. As shown in Fig. 5c (track 3), the 7.2-kb fragment is absent in the DNA isolated from strain SP488 (*mat1* inverted duplication). By this assay, as little as 1% of cells having breaks at both *smi* sites could have been readily detected. In the following section we show that the observed level of break at either *smi* site is similar to that found in strains containing only one *mat1* cassette, which argues against rapid repairing of one or the other *smi* break. The possibility that the 7.2-kb fragment is efficiently degraded *in vivo* and thus escapes detection is unlikely, as such strains do not generate dead or *Leu*⁻ progeny. In summary, in the direct duplication, both cassettes can become competent to switch in a given cell in the same cell cycle, while in the inverted arrangement, both cassettes cannot do so. As shown below, both cassettes of the inverted duplication are capable of acquiring the break efficiently. Thus, one daughter cell should get a break in one cassette and its sister should inherit a cut in the other cassette, as predicted in Fig. 5b by the strand segregation model.

To determine the efficiency with which each cassette acquires the break, the same sample of DNA that was used for Fig. 5c and isolated from SP487 (direct duplication) and SP488 (inverted duplication) strains was digested with *Hind*III endonuclease and analysed by Southern blotting. The various sizes of the restriction fragments predicted from *in vivo* cleavage of the *smi* sites (individually or together) in the *Hind*III-digested DNA are shown in Fig. 6a,b. The observed intensity of hybridization of various fragments to the *LEU2* probe demonstrates that the original *mat1* cassette as well as the inverted cassette are equally competent to switch (Fig. 6c, lane 3, compare the intensity of 9.1-kb and 12.2-kb bands). In an identical experiment I assayed the level of the double-stranded break by scanning the autoradiograph and found that 26% of DNA was cut in a wild-type strain which contains a single *mat1* cassette. A strain with the inverted duplication exhibited a similar high level of break both at the left (30%) and at the right (28%) *smi* site. Therefore, nearly 50% of the cells have a break at one or the other *smi* site. Thus, both cassettes in the inverted arrangement can be cut at a wild-type rate and cutting occurs independent of the cassette orientation. But the additional cassette in the direct duplication is cleaved with one-tenth to one-fifth the efficiency observed with the original *mat1* cassette (Fig. 6c, lane 2, compare intensity of the middle 11.1-kb band with the 9.1-kb band). In both constructions, the same 1,450-bp *Ssp*I fragment was used for producing duplications. The variation in the efficiency of cutting at *smi* sites may be due to differences in the precise location of the *smi* site in the chromatin (for instance, nucleosome phasing may affect cutting) or because a sequence at the junction of the insert was created in the direct duplication, reducing cutting of the inserted cassette in a *cis* manner. The 6.1-kb fragment in DNA of strain SP487 is not visible (Fig. 6c, lane 3) because the amount of this fragment is low (see above) but can be seen when the autoradiograph is exposed for a longer time. The 7.2-kb fragment is absent in the SP488 DNA (Fig. 6c, lane 3), confirming the results above.

To visualize all the possible fragments predicted in Fig. 6, the *Hind*III-digested DNA of strains SP487 and SP488 was also probed with the *mat1*-containing 10.4-kb fragment (Fig. 6c, lanes 4-6). In addition to the predicted bands (Fig. 6a,b), an 8.0-kb fragment was observed in the SP487 DNA (lane 4). This fragment is contributed by the *Leu*⁻ mitotic segregants (data not shown), which could have been generated by the loss of the 6.1-kb fragment and/or by switching events where one cassette is substituted for the duplication.

Discussion

These results show that the haploid cells can acquire a double-stranded break at more than one cassette in a given cell cycle in the directly oriented duplication. Thus both *mat1* cassettes in a haploid cell can acquire the competence to switch, consistent

with both the cytoplasmic and the strand segregation model. In contrast, although both cassettes are cleaved efficiently in the inverted duplication, only one or the other cassette is cleaved in a given cell. Thus, these results satisfy a key prediction (Fig. 5) of the strand segregation model and rule out the cytoplasmic segregation model. Also these results show that the site of imprinting must be contained within the 1,450-bp *mat1* sequence which was used for constructing duplications.

The fact that the total amount of break found in strains with the inverted duplication was twice the level found in standard strains which contain a single *mat1* cassette suggests that twice as many cells should switch in strains containing the inverted duplication. Thus both sister cells containing the inverted duplication are suggested to be capable of generating switched progeny in the next generation. A biological test of these predictions by pedigree analysis of cells undergoing switching is not at present possible, but the physical analysis presented allows us to conclude that the developmental asymmetry among sister cells must not be due to cytoplasmic differences but is a consequence of inheriting nonequivalent parental DNA chains.

I have examined the mechanism responsible for the localization of developmental information in the dividing cell. The results confirm and extend Egel's⁹ observation that switching potential segregates in *cis* with the *mat1* allele and suggest that an 'epigenetic' DNA strand-specific modification is catalysed at the *mat1* locus. The modification presumably allows the DNA molecule to be cut *in vivo* and switching follows. Such a hypothetical imprinting process is analogous to the restriction modification system of prokaryotes²⁴ except that in the yeast system DNA modification is a requirement for DNA cleavage. At present we can only speculate regarding the nature of the DNA modification. It may be DNA methylation or some other base modification, an unrepaired RNA primer of Okazaki fragments²⁵, a protein complex that segregates with a specific strand²⁶, or a site-specific single-stranded nick that becomes a double-stranded break in the next round of replication. Thus far no single-stranded nicks in the DNA have been found. The *swi1*, *swi3* and *swi7* gene functions are required for the inferred imprinting event (manuscript in preparation).

The fission yeast system is relatively simple because only two cell divisions are required to complete the developmental programme and produce two cell types. The developmental programme in multicellular organisms is clearly much more

complex. This work, however, suggests that DNA strands may provide a template for asymmetry among daughter cells in any biological system in any given cell division. For example, a modified strand or a preassembled transcriptionally competent complex may be delivered to only one daughter cell. A cascade of such controls may constitute a developmental programme required to generate a sequential pattern of cellular differentiation. For such a mechanism to operate, especially for higher systems, one has to further suppose that DNA chains are segregated to daughter cells nonrandomly. This constraint is needed because particular cell types are arranged geographically in a precise manner in the developing organism. Generation of a somatic and a reproductive cell by each of the 16 cells of *Volvox carter* at the 32-cell stage²⁷, pattern formation for heterocyst cell development in the blue-green algae *Anabaena catenula*²⁸, the ability of the *HO* gene to express in mother but not in bud cells of *Saccharomyces cerevisiae*^{29,30}, and many cases of cell-type determination in the cell lineage of *Caenorhabditis elegans*³¹ are just a few examples in which parental DNA strands may be segregated to daughter cells in a nonrandom fashion.

It should be noted that these results address only the cell-autonomous developmental cues. In multicellular organisms, the fate of certain cells is sometimes influenced by non-autonomous signals, for example, those specified by cell-cell interactions³². Such a mechanism is probably not required for the simpler system of unicellular fission yeast. It is worth stating, however, that the ability to respond to those extracellular signals in higher systems may be cell-autonomous and thus may be a consequence of inheriting different parental DNA chains by the sister cells.

I thank Barbara McClintock, James Watson, John Cairns, Frank Stahl, Richard Egel, Herbert Gutz, Terri Grodzicker, Donald Tipper, Steve Dellaporta, George Livi, Robert Cafferkey, Margaret Kelly and Doug Owen for discussions and/or help with the manuscript, Richard Egel and Mike Smith for communicating results before publication, James Garrels and Cecile Chang for helping me quantify the level of hybridization of various restriction fragments by scanning autoradiographs, Lisa Miglio for technical assistance, Mike Ockler, Jean Roberts and David Greene for preparing the figures and Laurie Lowman and Cathy Bosch for preparing the manuscript. This work is funded by a NIH grant and by the Cold Spring Harbor Laboratory Robertson Fund.

Received 21 October 1986; accepted 11 February 1987.

1. Gurdon, J. B. *J. Embryol. exp. Morph.* **10**, 622-640 (1962).
2. Holliday, R. & Pugh, J. E. *Science* **187**, 226-232 (1975).
3. Gutz, H., Heslot, H., Leupold, U. & Loprieno, N. in *Handbook of Genetics* (ed. King, R. C.) 395-446 (Plenum, New York, 1974).
4. Miyata, H. & Miyata, M. *J. gen. appl. Microbiol.* **27**, 365-371 (1981).
5. Egel, R. & Gutz, H. *Curr. Genet.* **3**, 5-12 (1981).
6. Beach, D. H. *Nature* **305**, 682-687 (1983).
7. Beach, D. H. & Klar, A. J. S. *EMBO J.* **3**, 603-610 (1984).
8. Egel, R. *Curr. Genet.* **8**, 199-203 (1984).
9. Egel, R. *Curr. Genet.* **8**, 205-210 (1984).
10. Crouse, H. V. *Genetics* **45**, 1429-1434 (1960).
11. Watson, J. D. & Crick, F. *Nature* **171**, 737-738 (1953).
12. Egel, R., Beach, D. H. & Klar, A. J. S. *Proc. natn. Acad. Sci. U.S.A.* **81**, 3481-3485 (1984).
13. Klar, A. J. S. & Miglio, L. M. *Cell* **46**, 725-731 (1986).
14. Piekawica, A., Bickle, T. A., Shepherd, J. C. W. & Ineichen, K. *J. molec. Biol.* **146**, 167-172 (1981).
15. Cairns, J. *Nature* **255**, 197-200 (1975).
16. Beach, D. H. & Nurse, P. M. *Nature* **290**, 140-142 (1981).
17. Orr-Weaver, T. L., Szostak, J. W. & Rothstein, R. J. *Proc. natn. Acad. Sci. U.S.A.* **78**, 6254-6358 (1981).

18. Klar, A. J. S., Hicks, J. B. & Strathern, J. N. *Cell* **28**, 551-561 (1982).
19. Rothstein, R. J. *Meth. Enzym.* **101**, 202-211 (1983).
20. Iino, Y. & Yamamoto, M. *Proc. natn. Acad. Sci. U.S.A.* **82**, 2447-2451 (1985).
21. Nurse, P. *Molec. gen. Genet.* **198**, 497-502 (1985).
22. Haber, J. E. & Rogers, D. T. *Nature* **296**, 768-770 (1982).
23. Southern, E. M. *Molec. gen. Genet.* **98**, 503-517 (1985).
24. Modrich, P. & Roberts, R. J. *Nucleases* (eds Lin, S. M. & Roberts, R. J.) 109-154 (Cold Spring Harbor Laboratory, New York, 1982).
25. Brutlag, D., Schekman, R. & Kornberg, A. *Proc. natn. Acad. Sci. U.S.A.* **68**, 2826-2830 (1971).
26. Brown, D. D. *Cell* **37**, 359-365 (1984).
27. Sumner, M. in *Pattern Formation* (eds Malacinski, G. M. & Bryant, S. V.) 197-212 (Macmillan, New York, 1984).
28. Wilcox, M., Mitchison, G. J. & Smith, R. J. *J. Cell Sci.* **12**, 707-711 (1973).
29. Nasmyth, K. *Nature* **302**, 670-676 (1983).
30. Jensen, R. E. & Herskowitz, I. *Cold Spring Harbor Symp. quant. Biol.* **49**, 97-104 (1984).
31. Sulston, J. E., Schierenberg, E., White, J. G. & Thomson, J. N. *Dev. Biol.* **100**, 64-119 (1983).
32. Sternberg, P. W. & Horvitz, H. R. *A. Rev. Genet.* **18**, 489-524 (1984).
33. Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor, New York, 1982).
34. Gutz, H. & Schmidt, H. *Curr. Genet.* **9**, 325-331 (1985).

The cell-cycle regulated proliferating cell nuclear antigen is required for SV40 DNA replication *in vitro*

Gregory Prelich, Matthew Kostura, Daniel R. Marshak, Michael B. Mathews & Bruce Stillman*

Cold Spring Harbor Laboratory, PO Box 100, Cold Spring Harbor, New York 11724, USA

Cell-free extracts prepared from human 293 cells, supplemented with purified SV40 large-T antigen, support replication of plasmids containing the SV40 origin of DNA replication. A cellular protein ($M_r \sim 36,000$) that is required for efficient SV40 DNA synthesis in vitro has been purified from these extracts. This protein is recognized by human autoantibodies and is identified as the cell-cycle regulated protein known as proliferating cell nuclear antigen (PCNA) or cyclin.

PREREQUISITE to understanding how cellular proliferation is controlled in eukaryotes is the identification of proteins that are directly involved in DNA replication, particularly those involved in its regulation. The transition into the replicative S phase of the cell cycle is accompanied by induction of a specific set of cellular proteins, some of which probably participate directly in DNA replication. A great deal has been learned about the enzymatic requirements for DNA replication in prokaryotes, largely because of the availability of well-defined phage and cellular origins of replication and of cell-free extracts that are capable of specifically replicating DNA molecules containing those origins¹⁻³. An analogous approach has been adopted in an effort to elucidate the mechanism of eukaryotic DNA replication. Little is known about the DNA sequences that constitute origins of replication in cellular chromosomes, so viral replicons have become the systems of choice. The development of soluble extracts capable of replicating adenovirus⁴, simian virus 40 (SV40)⁵⁻⁸ and polyomavirus⁹ DNAs *in vitro* has enabled rapid progress to be made in this field. The majority of the factors required for papovavirus DNA replication are of cellular origin and it is this dependence upon host proteins that makes these systems such attractive models to investigate the mechanism of cellular chromosomal replication.

SV40 contains a covalently closed DNA genome of 5,243 base pairs that is maintained in a chromatin structure similar to the structure of cellular chromosomes. Within the minichromosome lies a single origin of replication (*ori*) from which synthesis proceeds bi-directionally and semi-conservatively around the molecule¹⁰⁻¹². The only virus-encoded protein that is required for replication is the SV40 large tumour antigen (large-T antigen)¹³, a multifunctional phosphoprotein of relative molecular mass 90,000 (M_r 90K) which is necessary for a number of processes, including transcriptional control¹⁴⁻¹⁶ and cellular transformation¹⁷ as well as DNA replication¹⁸. Among its biochemical activities are its origin-specific DNA-binding property¹⁹ and an ATP-dependent DNA helicase function²⁰ that is associated with an ATPase activity. Mutant large-T antigens that are defective for either of these activities are replication defective²⁰⁻²³. Cytosolic extracts prepared from monkey or human cells support efficient SV40 DNA replication which is dependent upon purified large-T antigen and covalently closed circular plasmids containing the SV40 *ori* sequence, and closely resembles the mode of DNA synthesis observed in infected cells⁵⁻⁸. Addition of a nuclear extract enables the replicating DNA to be concomitantly assembled into minichromosomes²⁴. By fractionating the crude cellular cytosolic extract, we have purified a cellular factor that is required for SV40 DNA replica-

tion *in vitro* and have identified it as the proliferating cell nuclear antigen (PCNA)²⁵⁻²⁷, a protein recognized by antibodies from some patients with the autoimmune disease, systemic lupus erythematosus (SLE). PCNA, also known as cyclin²⁸, is a cell-cycle regulated protein that is expressed at elevated levels in transformed and tumour cell lines, and indirect studies have associated it with cellular DNA replication. Thus, this report establishes a direct biochemical link between a cell-cycle regulated, transformation-sensitive protein and the process of DNA replication. An accompanying paper shows that this factor stimulates the activity of the DNA polymerase- δ enzyme²⁹.

Cellular replication proteins

Our strategy for identification of the cellular replication factors is to divide the cytosolic extract derived from human 293 cells^{7,30} into multiple fractions, each of which will ultimately contain a single replication factor that can be purified. The post-ribosomal supernatant (S100) fraction was initially divided by phosphocellulose column chromatography into two components (fractions I and II, Fig. 1a). In standard reactions containing large-T antigen and origin-containing plasmid DNA, each of the these fractions alone was inactive, but efficient replication was restored when they were incubated together (Fig. 1b). Fraction I was further divided by DEAE-cellulose column chromatography into two components, A and B (Fig. 1a). The combination of fractions A and II gave a low level of DNA replication (Fig. 1b), which was dependent upon the presence of large-T antigen and a functional *ori* sequence as expected (data not shown and ref. 29). Addition of fraction B stimulated DNA synthesis greatly, restoring it to a level comparable to that given by the unfractionated cytosol S100 fraction.

We initially concentrated on fraction B because early results indicated that this fraction contains a single, essential replication component. Its purification was followed by its ability to stimulate replication in the presence of optimal amounts of fractions II and A, as measured by the incorporation of ³²P-dAMP into acid-precipitable material. Using this assay, the stimulatory activity present in fraction B was isolated by sequential phenyl-Sepharose column chromatography, Mono Q column chromatography and glycerol gradient velocity sedimentation (Table 1). In other experiments, the replication factor failed to bind to native and denatured DNA cellulose columns under a variety of conditions (data not shown). The purified protein obtained from a preparative glycerol gradient was diluted and re-run on an analytical glycerol gradient (Fig. 2). Fractions from the gradient were examined by SDS-polyacrylamide gel electrophoresis (PAGE) followed by staining with Coomassie brilliant blue dye (Fig. 2a), and were assayed for replication activity (Fig. 2b). The peak of replication activity corresponds exactly

* To whom correspondence should be addressed.

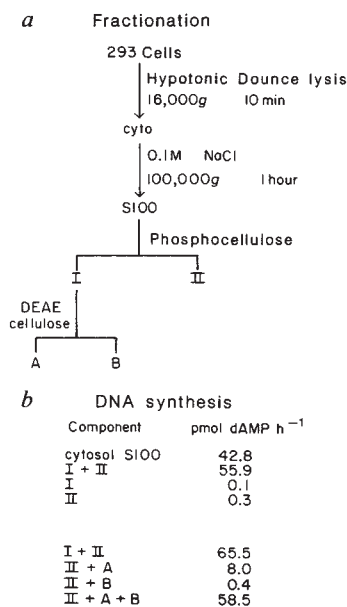


Fig. 1 *a*, Fractionation scheme of the 293 cytosol replication extract. The cytoplasmic S100 was derived from 293 cells as described^{7,30}. Briefly, cells were lysed by Dounce homogenization in a hypotonic buffer, and nuclei were removed by low-speed centrifugation. The supernatant was adjusted to 0.1 M NaCl and centrifuged at 100,000g for 1 h at 4 °C. The supernatant from this step is designated the S100. The S100 was fractionated by phosphocellulose and DEAE-cellulose chromatography (see Table 1 legend for details), resulting in fractions II, A and B, each of which is required for efficient replication. *b*, Replication activity of the fractionated system. Optimal amounts of individual fractions or the indicated combination of fractions were incubated under standard conditions⁷ (30 mM HEPES pH 7.5, 7 mM MgCl₂, 0.5 mM DTT, 100 µM each dCTP, dGTP, dTTP, 25 µM [α -³²P]dATP, 4 mM ATP, 200 µM each CTP, GTP, UTP, 40 mM creatine phosphate, 1 µg creatine phosphokinase, 0.3 µg pSV40 DNA, and 1 µg of SV40 large-T antigen) in 50 µl reactions for 1 h at 37 °C. Reactions were terminated by the addition of EDTA to 10 mM, precipitated by addition of 8% trichloroacetic acid (TCA) containing 1% sodium pyrophosphate and filtered through glass fibre filters (Enzo Biochem). Radioactivity was determined by liquid scintillation counting.

with the presence of an apparently homogeneous 36K protein. This protein co-fractionated with the replication activity through all chromatographic steps examined and restored full-length template replication (data not shown). Its sedimentation behaviour is consistent with the possibility that the protein is active as a homodimer, as it sediments slightly faster than the 66K bovine serum albumin (BSA) marker (5S).

PCNA/cyclin and DNA replication

While this work was in progress, other studies were addressing the nature and function of the proliferating cell nuclear antigen. PCNA was first identified as an antigen reactive with sera from some patients with SLE²⁵⁻²⁷. These sera yield weak or undetectable immunofluorescence in resting cells or normal tissue but give bright nuclear fluorescence with proliferating cells, including tumour cells, where they may have diagnostic value^{31,32}. The autoantibodies immunoprecipitate an acidic 36K protein that is identical to the cell-cycle regulated protein, cyclin²⁸. Cyclin was identified by two-dimensional gel electrophoresis as a transformation-sensitive protein that is preferentially synthesized during the S phase of the cell cycle³³ and subsequent work showed that its synthesis precedes replication by a short interval^{34,35}. In light of the accumulating circumstantial evidence linking PCNA/cyclin with DNA replication and/or cell-cycle progression through

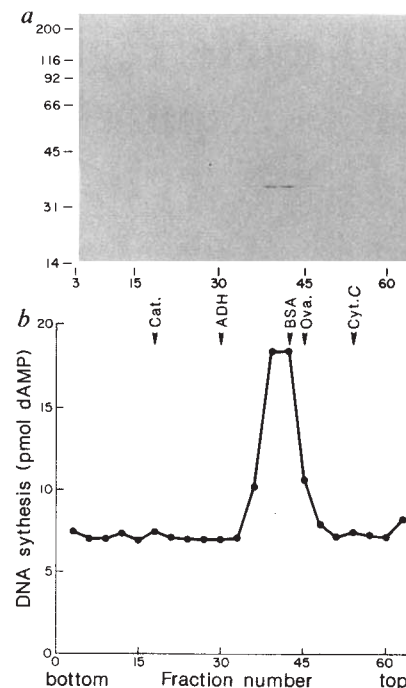
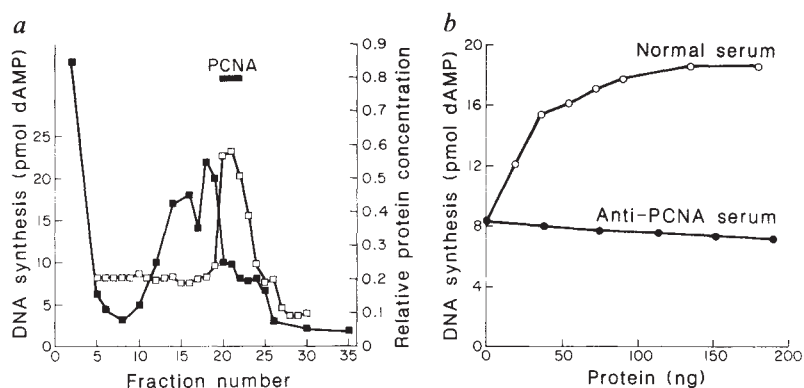


Fig. 2 *a*, Glycerol gradient sedimentation of the purified SV40 replication factor. A sample of the glycerol gradient purified replication factor (step 6, Table 1) was applied to a 15–35% glycerol gradient in buffer A containing 250 mM NaCl and centrifuged under conditions described in Table 1, legend. Every third fraction from the glycerol gradient was subjected to 10% SDS-PAGE. The gel was then stained with Coomassie brilliant blue dye. Markers at left, *M_r* in thousands, are cytochrome *c* (14K), carbonic anhydrase (31K), ovalbumin (45K), bovine serum albumin (66K), phosphorylase B (92K), β -galactosidase (116K) and myosin (200K). *b*, Every third fraction from the glycerol gradient was assayed for stimulation of replication in the presence of optimal amounts of fraction II and A, as in Fig. 1. Reactions were incubated under standard conditions at 37 °C for 2 h and were then terminated by adding EDTA to 10 mM, followed by precipitation in 8% TCA containing sodium pyrophosphate and filtration through glass-fibre filters. Radioactivity was determined by liquid scintillation counting. Protein markers sedimented in a parallel gradient were catalase (Cat); alcohol dehydrogenase (ADH); bovine serum albumin (BSA); ovalbumin (Ova) and cytochrome *c* (Cyt *c*).

S phase, we had begun to purify PCNA and study its function. PCNA was detected by immunoprecipitation in crude cytosol extracts isolated from human HeLa cells and was partially purified by ammonium sulphate fractionation, phenyl-Sepharose and DEAE-cellulose column chromatography. Like the SV40 replication factor, PCNA did not bind to either native or denatured DNA-cellulose columns (data not shown). Their marked physical similarities suggested that PCNA and the SV40 DNA replication factor described above might be one and the same.

To test this possibility, we used human anti-PCNA antibodies in conjunction with the replication system. First, an ammonium sulphate fraction derived from a HeLa cell cytosol S100 fraction was subjected to chromatography over a Mono Q resin and column fractions were assayed for their ability to stimulate replication in reactions containing fractions II and A, in addition to large-T antigen. This replication activity coincided precisely with the presence of PCNA, which was detected by immunoprecipitation with anti-PCNA antibody and gel electrophoresis (Fig. 3*a*). Thus the replication factor and PCNA behave identically in this chromatographic system. Second, a partially purified HeLa cell fraction containing HeLa cell PCNA (equivalent to step 5 in Table 1; mono Q pool) which could substitute for fraction B in replication reactions, was depleted of

Fig. 3 Identification of the SV40 replication protein as PCNA. *a*, Co-elution of the replication activity with PCNA on an anion-exchange column. A cytosolic extract prepared from human HeLa cells was fractionated by ammonium sulphate precipitation and the fraction containing PCNA (40–70% ammonium sulphate precipitate; determined by immunoprecipitation using human autoantibodies) was dialysed and chromatographed over a Mono Q column essentially as described in Table 1. ■, Relative protein content⁴⁵ of the column fractions; □, ability of each fraction to stimulate SV40 DNA replication when added to reactions containing fractions II and A. The peak of replication activity corresponded to a peak of the 36K protein detected by SDS-PAGE and was shown to be PCNA by immunoprecipitation using the human anti-PCNA autoantibodies (black bar). *b*, Depletion of the SV40 replication factor with anti-PCNA autoantiserum. A separate partially purified fraction containing HeLa cell PCNA (equivalent to the mono Q fraction, Table 1) was passed over Sepharose columns containing covalently bound immunoglobulin G (IgG) (5–7 mg ml⁻¹ bed volume) from either normal human serum or from an SLE patient's serum that contained anti-PCNA antibodies. The unbound fraction from each column was then added to replication reactions containing the cellular fractions II and A, and other components as indicated in Fig. 1 legend. Reactions were incubated for 1 h at 37 °C, and DNA synthesis was measured by incorporation of ³²P-dAMP into acid-insoluble material. The graph shows the activity of the HeLa PCNA-containing fraction after it was passed over a column containing normal human sera IgG (○) or anti-PCNA IgG (●).



PCNA by chromatography over a Sepharose column containing covalently bound antibody from normal human serum or antibody from an SLE patient²⁸. The protein that flowed through the column was assayed in replication reactions containing large-T antigen and fractions II and A derived from human 293 cells. Figure 3b shows that the unbound fraction from a column containing normal human antibody could still stimulate SV40 DNA replication *in vitro*, but that anti-PCNA antibody removed the active component. This establishes that the anti-PCNA antibodies react, either directly or indirectly, with a factor that stimulates DNA replication.

Protein sequence

The results above show that the replication factor and PCNA have the same apparent subunit M_r , exhibit similar chromato-

graphic behaviour, and are antigenically related. Furthermore, the purified HeLa cell protein, which was active in replication reactions, comigrated with PCNA during two-dimensional gel electrophoresis (data not shown). To establish their identity conclusively, we performed direct amino-acid sequence analysis of the two proteins. PCNA was purified from human HeLa cells or a commercial rabbit thymus extract by preparative immunoprecipitation, gel electrophoresis and electroelution, and was subjected to direct amino-acid sequencing using an Applied Biosystems Model 470A sequencer. The purified 36K SV40 replication factor isolated from human 293 cells was analysed directly. In all three cases, the following identical 26-residue sequence was obtained: H₂N-Met-Phe-Glu-Ala-Arg-Leu-Val-Gln-Gly-Ser-Ile-Leu-Lys-Lys-Val-Leu-Glu-Ala-Leu-Lys-Asp-Leu-Ile-Asn-Glu-Ala. This sequence is the same as the

Table 1 Purification of a replication component from fraction B

Fraction	Volume (ml)	Protein concentration (mg ml ⁻¹)	Total protein (mg)	Total units replication activity	Specific activity (units mg ⁻¹)	Recovery (%)	Purification factor
1 S100	93.0	13.8	1,278.7	—	—	—	—
2 I	21.2	29.5	625.4	—	—	—	—
3 B	5.0	6.8	33.8	425,000	12,925	100	1
4 Phenyl-Sepharose	2.6	1.4	3.5	539,130	153,162	127	12
5 Mono Q	1.1	0.33	0.36	224,639	623,998	53	49
6 Glycerol gradient	23.9	0.002	0.048	159,435	3,321,562	38	257

Protein concentration was determined by the method of Bradford⁴⁵ using BSA as a standard. One unit of replication activity incorporated 1 pmol of ³²P-dATP into acid-precipitable material in 1 h at 37 °C in the presence of saturating amounts of fractions II and A.

Methods. The cytosol replication extract was derived from 32 l of suspension 293 cells as described, then adjusted to 100 mM NaCl and centrifuged at 100,000g for 1 h at 4 °C^{7,30}. The supernatant fraction, designated the S100 cytosol fraction, was then adjusted to 0.2 M NaCl and loaded onto a 150-ml phosphocellulose column that had been equilibrated in buffer A (25 mM Tris-HCl pH 7.5, 1 mM EDTA, 0.01% Nonidet P40, 10% glycerol, 1 mM dithiothreitol (DTT), 0.1 mM phenylmethylsulphonyl fluoride (PMSF) containing 0.2 M NaCl). The column was washed with 300 ml of buffer A containing 0.2 M NaCl and protein was eluted with buffer A containing 1 M NaCl. The peak of protein was determined by the method of Bradford⁴⁵. The unbound peak (fraction I) and the bound peak (fraction II) were then dialysed extensively against buffer A containing 25 mM NaCl and 20% sucrose. Fraction I was adjusted to 175 mM NaCl and applied to a 200 ml DEAE-cellulose (Whatman DE-52) column that had been equilibrated with buffer A containing 175 mM NaCl, and protein was eluted with buffer A containing 1 M NaCl. The peak of protein was determined, and the unbound (fraction A) and bound (fraction B) peaks were pooled and dialysed against buffer A containing 25 mM NaCl and 20% sucrose. Fraction B was then adjusted to 1 M (NH₄)₂SO₄ and applied to a 20-ml phenyl-Sepharose (Pharmacia) column that had been equilibrated in the same buffer. The column was then washed successively with 100 ml of buffer A containing 1 M (NH₄)₂SO₄, 100 ml of buffer A containing 0.5 M (NH₄)₂SO₄ followed by a gradient from this buffer to buffer A without (NH₄)₂SO₄. The fractions were tested for stimulation of replication activity using optimal concentrations of fractions II and A. Fractions containing a single peak of stimulatory activity that eluted toward the end of the gradient were pooled and dialysed against buffer A containing 25 mM NaCl and 20% sucrose. The sample was adjusted to 0.2 M NaCl, and loaded onto a 1-ml Mono Q HR5/5 column (Pharmacia) equilibrated with buffer A containing 0.2 M NaCl pH 8.0. The column was washed with 2 ml of buffer A (pH 8.0) containing 0.2 M NaCl and eluted with an 8-ml gradient of buffer A (pH 8.0) containing 0.2 M NaCl to buffer A (pH 8.0) containing 0.6 M NaCl. The fractions were assayed for replication activity, and peak fractions which eluted at 400 mM NaCl were pooled. Pooled fractions (100 μl) were applied to a 15–35% glycerol gradient in buffer A (pH 7.5) containing 250 mM NaCl. The gradient was centrifuged at 4 °C at 49,000 r.p.m. for 21 h in a Beckman SW 50.1 rotor. Fractions were collected from the bottom and those fractions containing replication activity were pooled, aliquoted and stored at -70 °C. The glycerol gradient step was performed on an analytical scale only, but the Table shows the total activity expected.

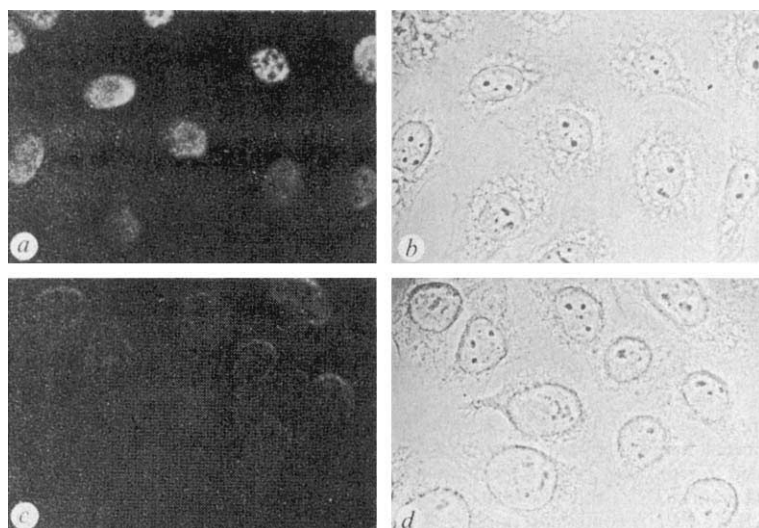


Fig. 4 Blocking of PCNA nuclear immunofluorescence by the 36K replication factor. Glass coverslips containing asynchronously growing HeLa cells were washed for 5 min with phosphate-buffered saline (PBS), fixed for 10 min with methanol at -20°C , and washed for 5 min with PBS. They were then incubated for 30 min at room temperature with anti-PCNA autoantiserum²⁸ that had been incubated for 1 h on ice with either 1 μg of bovine serum albumin (a and b) or 0.5 μg of the purified 36K replication factor (c and d). They were then washed extensively with PBS before being incubated for 30 min at room temperature in the presence of fluorescein-conjugated goat anti-human antibodies (Cappel Laboratories). The coverslips then were washed with PBS and mounted with Immomount (Shandon Southern Industries). Immunofluorescence (a and c) and phase-contrast micrographs (b and d) of the same field are shown.

amino-terminal sequence of rabbit PCNA presented previously³⁶, except that the earlier authors did not obtain a definitive assignment of the tenth amino acid as serine. The sequence identity convincingly demonstrates that the 36K replication protein is structurally related to immunoprecipitated PCNA, and taken together with their similar physical and biological properties, leaves no doubt that PCNA is required for efficient SV40 DNA replication *in vitro*.

Does PCNA act in DNA synthesis *in vivo*?

PCNA was initially defined on the basis of characteristic nuclear immunofluorescence patterns obtained when proliferating cells are stained with autoantibodies from a small subset of SLE patients²⁵⁻²⁷. In non-synchronous cell populations a variety of patterns is observed, which in synchronized cultures appear to succeed one another in an orderly way through the cell cycle^{34,35,37,38}. The sites of immunofluorescence correspond to the sites of active DNA synthesis visualized by ^3H -thymidine incorporation and emulsion autoradiography, suggesting that PCNA is associated with DNA synthesis in the cell^{34,37,39}. Unfortunately, however, autoimmune sera are rarely monospecific and often contain antibodies directed against other nuclear components such as DNA, histones and ribonucleoprotein particles, which might confuse the issue. Indeed, the human sera used in many studies recognize multiple denatured proteins in immuno-blots, but only PCNA by immunoprecipitation⁴⁰. To determine whether the nuclear fluorescence observed in proliferating cells is due to PCNA, we studied the ability of the purified replication factor to quench the immunofluorescence.

Non-synchronized human HeLa cells grown at low density on glass coverslips were fixed and stained with anti-PCNA autoantiserum, followed by fluorescein-labelled goat anti-human immunoglobulin G (IgG) second antibody (Fig. 4a). Characteristic nuclear fluorescence was observed that had a heterogeneous staining pattern. When the antibody was pre-incubated with glycerol-gradient-purified 36K SV40 replication factor (PCNA), all nuclear-staining patterns were eliminated, leaving only a weak, peripheral nuclear fluorescence that was observed with normal human serum (Fig. 4c). Thus, the replication protein is able to block the nuclear fluorescence staining pattern that first defined PCNA. We conclude that the nuclear antigen observed by immunofluorescence and associated with cells in the S phase of the cell cycle is related to the 36K replication protein, and infer that this factor indeed participates in cellular DNA synthesis *in vivo*.

Function of PCNA

To examine the stage of replication at which PCNA acts, a time course of replication was conducted using the fractionated sys-

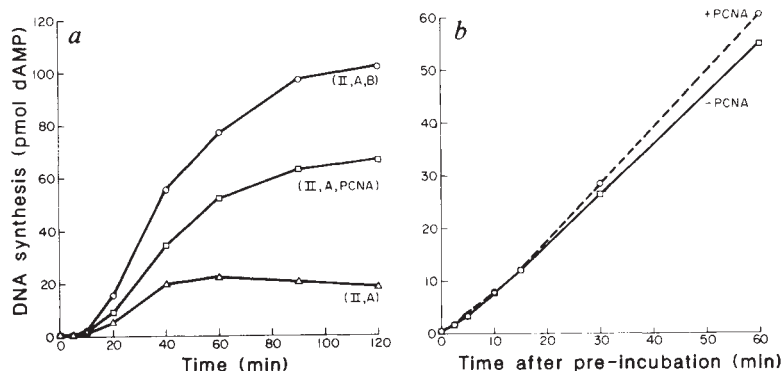
tem in the presence or absence of either fraction B or the 36K protein. As shown in Fig. 5a, the kinetics of replication using the complete fractionated system (fractions II, A and B) are similar to results obtained previously with the S100 extract^{7,30}. Very little synthesis occurs during the first 10 min of incubation, but extensive incorporation follows over the next two hours. In the absence of fraction B, the initial 10-minute lag was still observed, followed by a short burst of DNA synthesis, but then no subsequent increase in incorporation occurred over the remainder of the time course. Addition of PCNA to reactions containing fractions II and A resulted in the re-appearance of normal kinetics of DNA replication, although the absolute level of replication was less than that obtained using the more crude fraction B. There may be an additional stimulatory factor(s) present in fraction B, but it is not essential for complete replication (Fig. 5 and data not shown).

The initial lag in DNA synthesis is eliminated by pre-incubating the template DNA with the S100 extract and large-T antigen at 37°C in the absence of deoxynucleoside triphosphate (dNTP) precursors^{30,41}. During the preincubation, a stable pre-synthesis complex is formed on the template DNA that promotes immediate initiation of DNA replication on addition of dNTPs. This observation offers an assay for the identification of cellular proteins that are required for the pre-synthesis step in replication. To see if PCNA is involved in this stage, the template DNA was pre-incubated with large-T antigen and cellular fractions II and A in the presence or absence of PCNA. Upon addition of dNTPs (and of PCNA when absent from the pre-incubation), immediate DNA replication was observed, irrespective of whether or not PCNA was present in the pre-incubation reaction (Fig. 5b). However, by a similar analysis, both fractions II and A were required for the pre-synthesis complex formation (data not shown), suggesting that they contain factors required for initiation. These results indicate that PCNA is not required for the formation of the pre-synthesis complex on the template DNA, but functions at a subsequent stage, presumably elongation.

Discussion

PCNA/cyclin was originally described on the basis of studies with autoantibodies and with two-dimensional gels, as a nuclear antigen that is present in proliferating cells and in elevated amounts in transformed cell lines and tumour cells (reviewed in refs 35 and 42). The increased PCNA content of tumorigenic cells most probably reflects the fact that they are proliferating rapidly, having lost normal cell growth control. These observations led to speculation that PCNA might be an intracellular component of cellular growth regulatory pathways and, more specifically, a protein involved in cellular DNA replication^{35,42}. Circumstantial support for an association with cellular DNA

Fig. 5 Kinetics of DNA replication in the presence and absence of fraction B or the purified 36K replication factor. *a*, Reactions (100 μ l) were set up on ice in the presence of the indicated cellular-derived fractions under the standard conditions of Fig. 1. The reactions then were incubated at 37 °C, and at the times shown aliquots were removed, precipitated with TCA, and the radioactivity determined by liquid scintillation counting. *b*, Reactions (100 μ l) were set up on ice containing 30 mM HEPES pH 7.5, 7 mM MgCl₂, 0.5 mM DTT, 200 μ M each CTP, GTP, UTP, 40 mM creatine phosphate, 2 μ g creatine phosphokinase, 0.6 μ g pSV40 DNA, and 2 μ g large-T antigen with optimal amounts of fractions II and A (minus deoxynucleoside triphosphates). One reaction (broken line) also contained 300 ng of PCNA (Table 1, step 6). The reactions were pre-incubated at 37 °C for 15 min, placed on ice, and 300 ng of PCNA was added to the reaction that lacked PCNA in the preincubation. Then dNTPs (100 μ M each dCTP, dGTP, dTTP and 25 μ M [α -³²P]dATP) were added to both reactions, and the DNA synthesis reaction was performed at 37 °C. At the indicated times, aliquots were removed, precipitated with TCA, and radioactivity determined. The presence or absence of PCNA in the pre-synthesis step is indicated.



replication was drawn from the observation that the nuclear immunofluorescent patterns in single cells roughly correspond to the sites of DNA synthesis detected by *in situ* autoradiography of ³H-thymidine-labelled cells^{34,37,39,42}. Our demonstration that PCNA is a cellular component required for SV40 DNA replication *in vitro*, and that it is responsible for the nuclear fluorescence pattern, argues strongly that this protein is involved in replication of cell chromosomes.

Cellular DNA replication is induced by a number of DNA tumour viruses, including adenovirus and SV40¹⁷ and it has recently been demonstrated⁴³ that the adenovirus *E1A* gene activates PCNA synthesis concomitant with induction of cellular DNA replication. We therefore expect that SV40 will also activate PCNA synthesis during infection, before the onset of SV40 DNA replication. PCNA synthesis is very low in quiescent, uninfected cells, but increases 6–7-fold just before S phase in cells treated with a variety of mitogenic agents^{34,35}. The pattern of PCNA immunofluorescence varies throughout the S phase of the cell cycle, but is reduced or undetectable in cells in the G₁ and G₂ phases and during mitosis^{34,35,37–39,42}. Despite the cell-cycle dependence of both PCNA synthesis and immunofluorescence, the protein is stable and can be detected by gel electrophoresis at all phases of the cell cycle³⁴. It seems possible, therefore, that DNA synthesis requires newly synthesized PCNA, or that its antigenic epitope is only exposed while the protein is engaged in its replication function. PCNA synthesis may indeed trigger replication. Towards the end of S phase, PCNA synthesis normally subsides as DNA replication declines, but in the presence of drugs that inhibit DNA replication, PCNA synthesis continues and only subsides on removal of the drug and after DNA replication has taken place³⁴. This raises the

possibility that PCNA synthesis is shut off by the act of DNA replication, perhaps by replication of the PCNA gene itself.

Further studies are needed to determine the precise role of PCNA in replication, but the available evidence suggests that it functions during the elongation stage of DNA replication. In an accompanying paper²⁹, we show that PCNA is identical with a DNA polymerase- δ accessory protein that enables this DNA polymerase to efficiently utilize template/primers containing long stretches of single-strand template⁴⁴, a function much like that of the β subunit of the *Escherichia coli* DNA polymerase III holoenzyme in replication processivity¹. This observation suggests a function for DNA polymerase- δ in replicative DNA synthesis and prompts a reinvestigation of the polymerases involved. Clearly, characterization of the cellular proteins that are involved in SV40 DNA replication should continue to give valuable insight into the mechanism of DNA replication in mammalian cells. In the present study, this approach has led to the identification of one such replication factor as the proliferating cell nuclear antigen or cyclin, a protein implicated in cell cycle control and chromosomal DNA replication.

We thank C. C. Bunn (Postgraduate Medical School, Hammersmith Hospital, London) and Dr R. M. Bernstein (Manchester Royal Infirmary) for serum samples, Stuart MacNeill for some initial studies on fractionation, B. Faha for technical assistance and B. Weinkauff for preparation of this manuscript. This research was supported by grants from the US National Cancer Institute (CA 13106) and from the Rita Allen Foundation (to B.S.). G.P. is a predoctoral fellow and was partially supported by training grant GM07964 and M.K. is a postdoctoral fellow supported by training grant CA09311, both from the US Public Health Service.

Received 17 November 1986; accepted 22 January 1987.

- Kornberg, A. *DNA Replication* (Freeman, San Francisco, 1980).
- Kornberg, A. *DNA Replication: 1982 Supplement* (Freeman, San Francisco, 1982).
- Marians, K. J. *CRC Crit. Rev. Biochem.* **17**, 153–215 (1984).
- Challberg, M. D. & Kelly, T. J. Jr *Proc. natn. Acad. Sci. U.S.A.* **76**, 655–659 (1979).
- Li, J. J. & Kelly, T. J. *Proc. natn. Acad. Sci. U.S.A.* **81**, 6973–6977 (1984).
- Li, J. J. & Kelly, T. J. *Molec. cell. Biol.* **5**, 1238–1246 (1985).
- Stillman, B. W. & Gluzman, Y. *Molec. cell. Biol.* **5**, 2051–2060 (1985).
- Wobbe, C. R., Dean, F. B., Weissbach, L. & Hurwitz, J. *Proc. natn. Acad. Sci. U.S.A.* **82**, 5710–5714 (1985).
- Murakami, Y., Eki, T., Yamada, M., Prives, C. & Hurwitz, J. *Proc. natn. Acad. Sci. U.S.A.* **83**, 6347–6351 (1986).
- DePamphilis, M. L. & Wassarman, P. M. in *Organization and Replication of Viral DNA* (ed. Kaplan, A. S.) 37–114 (CRC Press, Florida, 1982).
- Sundin, O. & Varshavsky, A. *Cell* **25**, 659–669 (1981).
- Weaver, D. T., Fields-Berry, S. C. & DePamphilis, M. L. *Cell* **41**, 565–575 (1985).
- Tegtmeyer, P. *J. Virol.* **10**, 591–598 (1972).
- Tegtmeyer, P., Schwartz, M., Collins, J. K. & Rundell, K. J. *J. Virol.* **16**, 168–178 (1975).
- Myers, R. M., Rio, D. C., Robbins, A. K. & Tjian, R. *Cell* **25**, 373–384 (1981).
- Hansen, U., Tenen, D. G., Livingston, D. M. & Sharp, P. A. *Cell* **27**, 603–612 (1981).
- Toozie, J. *Molecular Biology of Tumour Viruses* 2nd edn, part 2 (Cold Spring Harbor Laboratory, New York, 1980).
- Rigby, P. W. J. & Lane, D. P. *Adv. viral Oncol.* **3**, 31–57 (1983).
- Tjian, R. *Cell* **26**, 1–2 (1981).
- Stahl, H., Dröge, P. & Knippers, R. *EMBO J.* **5**, 1939–1944 (1986).
- Clark, R., Peden, K., Pipas, J., Nathans, D. & Tjian, R. *Molec. cell. Biol.* **3**, 220–228 (1983).
- Kalderon, D. & Smith, A. E. *Virology* **139**, 109–137 (1984).

- Stillman, B., Gerard, R. D., Guggenheimer, R. A. & Gluzman, Y. *EMBO J.* **4**, 2933–2939 (1985).
- Stillman, B. *Cell* **45**, 555–565 (1986).
- Miyachi, K., Fritzler, M. J. & Tan, E. M. *J. Immun.* **121**, 2228–2234 (1978).
- Takasaki, Y., Fishwild, D. & Tan, E. M. *J. exp. Med.* **159**, 981–992 (1984).
- Tan, E. M. *Adv. Immun.* **33**, 167–240 (1982).
- Mathews, M. B., Bernstein, R. M., Franza, B. R. Jr & Garrels, J. I. *Nature* **309**, 374–376 (1984).
- Prellich, G. *et al. Nature* **326**, 517–520 (1987).
- Stillman, B. W., Diffley, J. F. X., Prellich, G. & Guggenheimer, R. A. in *Cancer Cells 4 DNA Tumour Viruses: Control of Gene Expression and Replication* (eds Botchan, M., Grodzicker, T. & Sharp, P.) 453–463 (Cold Spring Harbor Laboratory, New York, 1986).
- Takasaki, Y., Robinson, W. A. & Tan, E. M. *J. natn. Cancer Inst.* **73**, 655–661 (1984).
- Smetana, K., Gyorkey, F., Chan, P. K., Tan, E. & Busch, H. *Blut* **46**, 133–141 (1983).
- Bravo, R. & Celis, J. E. *J. Cell Biol.* **84**, 795–802 (1980).
- Bravo, R. & McDonald-Bravo, H. *EMBO J.* **4**, 655–661 (1985).
- Bravo, R. *Exp. Cell Res.* **163**, 287–293 (1986).
- Ogata, K., Ogata, Y., Nakamura, R. M. & Tan, E. M. *J. Immun.* **135**, 2623–2627 (1985).
- Celis, J. E., & Celis, A. *Proc. natn. Acad. Sci. U.S.A.* **82**, 3262–3266 (1985).
- Kurki, P., Vanderlaan, M., Dolbeare, F., Gray, J. & Tan, E. M. *Exp. Cell Res.* **166**, 209–219 (1986).
- Madsen, P. & Celis, J. E. *FEBS Lett.* **193**, 5–11 (1985).
- Sadaie, M. R. & Mathews, M. B. *Exp. Cell Res.* **163**, 423–433 (1986).
- Wobbe, C. R., Dean, F. B., Murakami, Y., Weissbach, L. & Hurwitz, J. *Proc. natn. Acad. Sci. U.S.A.* **83**, 4612–4616 (1986).
- Celis, J. E., Madsen, P., Nielsen, S. & Celis, A. *Leukemia Res.* **10**, 237–249 (1986).
- Zerler, B., Roberts, R. J., Mathews, M. B. & Moran, E. *Molec. cell. Biol.* **7**, 821–829 (1987).
- Tan, C., Castillo, C., So, A. G. & Downey, K. M. *J. biol. Chem.* **261**, 12310–12316 (1986).
- Bradford, M. M. *Analyt. Biochem.* **72**, 248–254 (1976).

Upper limit on the mass of the electron neutrino

J. N. Bahcall* & S. L. Glashow†

The Institute for Advanced Study, Princeton, New Jersey, USA

† Harvard University, Cambridge, Massachusetts 02138, USA

The historic detection by the Kamiokande-II collaboration¹ and the IMB collaboration² of neutrinos from the Large Magellanic Cloud (LMC) supernova provides the first opportunity to determine the mass, m_{ν_e} , of the electron neutrino from astronomical observations. Here we show that m_{ν_e} is less than 11 eV, provided only that propagation effects have not conspired to sharpen, by more than a factor of two the narrow pulse-width of neutrinos, observed by the Kamiokande-II collaboration from the LMC supernova. This result improves on the laboratory limit on m_{ν_e} and confirms the view that electron neutrinos do not constitute the major component of the matter density of the Universe.

Many authors have discussed the fact that higher energy neutrinos from a distant supernova will arrive at Earth before lower energy neutrinos if the neutrino has a finite mass^{3,4}. For the recently observed LMC supernova, the additional travel time, Δt , that is required for a neutrino of energy E because it has a mass m_{ν_e} , where d is distance to the LMC:

$$\Delta t = 0.026 (d/50 \text{ kpc}) (m_{\nu_e}/1 \text{ eV})^2 (10 \text{ MeV}/E)^2 \quad (1)$$

For any given neutrino mass, one can calculate from equation (1) the emission time (offset by the light travel time) at the star for each neutrino that causes an observable event. The data contain no further information. How does one decide if the calculated departure times are 'acceptable'? Previous authors have not discussed this question.

Without some assumption about what is physically or astronomically plausible, one cannot proceed further. We do not have any *a priori* knowledge of the departure times of the neutrinos. Whatever the neutrino mass, the complicated nuclear and hydrodynamic processes occurring in the explosive formation of the neutron star might somehow yield a sequence of emission times in just such a way as to produce the sharply peaked time structure seen at Kamioka. Indeed, our analysis is based solely on the assumption that the narrow (in time) observed pulse is not an accident resulting from propagation effects.

Before describing our method, we will summarize what is known about the energies and arrival times of the observed Kamiokande-II events. In both these water detectors, the dominant events are $\bar{\nu}_e$ absorption by protons and ν_e scattering by electrons.

The first eight neutrino events arrived within less than 2 seconds, followed by a period of 7 seconds in which no neutrinos were detected. There were then three less certain neutrino events in the next 3 seconds. In the initial two-second pulse, five of the eight events were observed in the first 0.5 s. The events corresponded to measured positron (or electron) energies ranging from 7.5 MeV to 35 MeV, with typical uncertainties of the order of 20%. (In deriving mass limits, we have converted the observed (positron or electron) energies to neutrino energies by supposing that each event was due to absorption on protons by antineutrinos. The IMB events reported so far are all relatively high in energy (≥ 20 MeV) and are therefore less useful in setting limits on the neutrino mass.

If the observed pulse duration were caused solely by the mass of the neutrino, then the higher energy events would have arrived first. This was not observed. The time-ordered sequence of positron (or electron) energies was (all in MeV) 20, 13.5, 7.5, 9.2, 12.8, 35.4, 21.0, 19.8, 8.6, 13.0 and 8.9. Two of the lowest energy events (7.5 MeV and 9.2 MeV) preceded by 1.2 s three

Table 1 The width of the neutrino pulse at the neutron star

m_{ν_e} (eV)	Δ_8 (seconds)	Δ_{11} (seconds)
0.0	1.9	12.4
5.0	2.3	12.4
9.0	3.8	12.8
10.0	4.4	13.0
15.0	8.3	14.8
20.0	13.8	18.4
26.6	23.6	24.9

The maximum difference in departure times of the first eight events is Δ_8 and of the entire eleven events is Δ_{11} .

higher energy events (35 MeV, 21 MeV and 19.8 MeV). We conclude that the duration of the pulse at the source was at least 1.2 s and could have been (for zero neutrino mass) as long as the entire observed pulse.

The only assumption that we will make is that propagation effects did not sharpen the pulse width of the observed neutrino events by more than a factor of two. Thus the eight events which were detected at Earth in the initial two seconds of the pulse are presumed not to have been spread out at the source by more than four seconds. The choice of a factor of two is arbitrary but we believe that it is conservative and, if the reader prefers a different numerical factor, he can make the appropriate scaling using Table 1 (above). Stated differently, we assume that the effects of propagation on the arrival times have not been such as to produce by accident a short pulse from what was initially a much more diffuse signal.

Table 1 shows as a function of assumed neutrino mass the computed duration, at the neutron star, of both the first eight events (Δ_8) and all the reported events (Δ_{11}). For $m_{\nu_e} = 0$, these quantities are equal to the observed values of 2 s and 13 s. The inferred durations at the star increase monotonically with assumed mass because the observed events do not have the relation between energy and arrival time that is implied by equation (1). Using the best estimates given by the Kamiokande collaboration for all of the energies we find that if m_{ν_e} exceeds 9.0 eV, then the 2 s initial pulse that is observed on Earth would have been longer than 4 s at the neutron star. We have also recalculated the upper limit by varying the nominal energies within their quoted errors. In all cases, the pulse width at the source is at least twice the value measured at Earth if $m_{\nu_e} \geq 11$ eV. We conclude therefore that

$$m_{\nu_e} \leq 11 \text{ eV} \quad (2)$$

The strength of this limit derives from the lower energy events (7.5 MeV, 9.2 MeV and 12.8 MeV), all of which are observed to move at large angles with respect to the direction of the LMC (108, 70 and 135 degrees respectively). The measured directions show that these three events are not due to electron-neutrino scattering, which is strongly forward peaked.

A neutrino mass of 27 eV is required, according to Table 1, to make the entire set of eleven observed events twice as long (25 s) at the source than is observed at Earth. We reject this possibility as physically implausible because it implies that the initial pulse (observed to be 1.9 s in duration at the Earth, and surely a physically significant phenomenon) was compressed in time by more than an order of magnitude (from 22.4 s at the star).

There are other ways, most of them more complicated, of analysing the Kamiokande-II data. One can, of course, obtain more stringent limits provided one is willing to introduce stronger assumptions regarding what is astronomically or physically plausible. For example, if one could justify the claim that the five events observed in the first 0.5 seconds had a distinct astrophysical origin, one could lower the upper limit to 7.5 eV including the measurement errors.

Conservatively, we conclude that the mass of the electron neutrino must be less than or equal to 11 eV. This limit is stronger than has been achieved by 50 years of measurements on terrestrial systems.

Larger detectors sensitive to lower energy neutrinos should be able to observe, with much higher event rates, subsequent supernovae explosions in the Galaxy. Supernovae in the Galaxy could be as frequent as one a decade⁶ if they occur in obscured regions and are therefore not associated with observable optical supernovae. The next observation of neutrinos from a supernova explosion will provide an independent measurement of the neutrino mass and will exclude the extremely unlikely possibility that the LMC distance, the detected neutrino energies and the departure times all conspired to sharpen greatly the neutrino signal from the collapse. With larger detectors, it should be possible to set even better limits on (or measure) the mass of the electron neutrino. The higher event rates will permit experimentalists to determine characteristics of the ambient material at the time the neutron star is being formed. In the extended periods between observable explosions, the detectors can be used to measure accurately the arrival times, energies and directions of solar neutrinos.

We thank the Kamiokande-II collaboration and the IMB collaboration for sharing with us their epochal data. This work was supported in part by the NSF.

Received Friday 13 March; accepted 23 March 1987.

1. Hirata, K. *et al.* *Phys. Rev. Lett.* (submitted).
2. Bionata, R. M. *Phys. Rev. Lett.* (submitted).
3. Cabibbo, N. *Proc. Int. Meet. Astrophysics and Elementary Particles, Common Problems*, Rome, Italy (eds Cabibbo, N., Dallaporta, N., Gratton, L., Pacini, F. & Salvini, G.) (Accademia Nazionale dei Lincei, Rome, 1980).
4. Piran, T. *Phys. Lett. B* **102**, 299–302 (1981).
5. Bahcall, J., Dar, A. & Piran, A. *Nature* **326**, 135–136 (1987).
6. Bahcall, J. N. & Piran, T. *Astrophys. J.* **167**, L77–L81 (1983).

Unusual stable isotope ratios in amino acid and carboxylic acid extracts from the Murchison meteorite

S. Epstein*, R. V. Krishnamurthy*, J. R. Cronin†, S. Pizzarello† & G. U. Yuen†

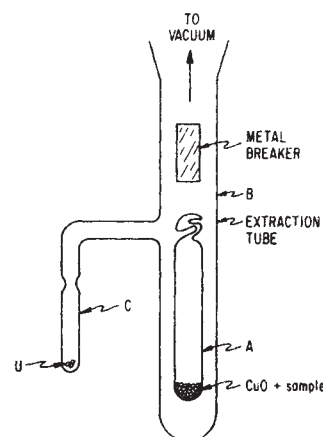
* Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, California 91125, USA

† Department of Chemistry and Center for Meteorite Studies, Arizona State University, Tempe, Arizona 85287, USA

Much effort has been directed to analyses of organic compounds in carbonaceous chondrites because of their implications for organic chemical evolution and the origin of life. We have determined the isotopic composition of hydrogen, nitrogen and carbon in amino acid and monocarboxylic acid extracts from the Murchison meteorite. The unusually high D/H and ¹⁵N/¹⁴N ratios in the amino acid fraction ($\delta D = 1,370\%$ after correction for isotope exchange; $\delta^{15}N = 90$) are uniquely characteristic of known interstellar organic materials. The δD value of the monocarboxylic acid fraction is lower (377%), but still consistent with an interstellar origin. These results confirm the extraterrestrial origin of both classes of compound, and provide the first evidence suggesting a direct relationship between the massive organo-synthesis occurring in interstellar clouds and the presence of pre-biotic compounds in primitive planetary bodies. The isotope data also bear on the historical problem of distinguishing indigenous material from terrestrial contaminants.^{1,2}

There have been previous investigations of the isotopic composition of soluble organic compounds from carbonaceous

Fig. 1 Apparatus, made from Vycor, for the isotope analysis of the samples. It was cleansed in hot Chromerge, triple quartz distilled water and preheated to 900 °C to dry and remove any traces of organic matter. When appropriate, the apparatus is torched under vacuum to remove possible terrestrial H₂O contamination.



chondrites. Smith and Kaplan³ and Chang and Mack⁴ measured $\delta^{13}C$ values of -17 to -5% for benzene-methanol and 5 to 25% for water extracts respectively. Chang and Mack also examined amino acids from the Murchison meteorite and obtained $\delta^{13}C$ values as high as 44% , which is more positive than most terrestrial organic material. Except for a few measurements of more than 40% , values of $\delta^{13}C$ are generally very similar to those of terrestrial carbon. Becker and Epstein⁵ analysed the isotopic composition of C, H and N of methanol and carbon tetrachloride extracts. They found nitrogen with $\delta^{15}N$ of 87% and hydrogen with δD as high as 486% for some of the methanol extracts. They suggested that amino acids might be present in the MeOH extract however, because hydrogen and nitrogen of HF-HCl insoluble residues had given δD values over $2,000\%$ and $\delta^{15}N$ as high as 150% ^{6–9}, it could not be assumed that the high δD and $\delta^{15}N$ of the methanol extract were due to amino acids.

In this work, well-characterized suites of both amino acids and carboxylic acids were obtained from the water-soluble fraction of the Murchison meteorite and subjected to isotope analyses. A 7.2 g interior piece of the meteorite was pulverized and extracted with water in a sealed, evacuated vial at $110^\circ C$. After centrifugation, the extract was decanted from the insoluble residue and acidified with hydrochloric acid. All volatile components (neutral and acidic) were distilled under vacuum and trapped at liquid nitrogen temperature. The distillate was adjusted to pH 13 and oven-dried at $100^\circ C$ to remove neutral species. The resulting acid salts were redissolved in water, acidified with sulphuric acid and again vacuum distilled and trapped at liquid nitrogen temperature. Gas-chromatography (GC) analysis of this distillate showed only the expected carboxylic acid peaks. The preparation was adjusted to pH 9–9.5 and dried.

Amino acids were isolated from the involatile residue left after the first vacuum distillation (see above). This residue was redissolved in 6N HCl and hydrolyzed in an evacuated vial for 24 h at $110^\circ C$. After drying in a vacuum desiccator, the sample was redissolved in 0.1N HCl and applied to a cation exchange column (BioRad AG-50W) which was then washed thoroughly with 6N HCl and with water to remove neutral and acidic species. Amino acids and amines were eluted with 2N NH₄OH. The sample was dried repeatedly on a rotary evaporator to remove any volatile amines.

The individual components of both preparations are well known from previous work. Fifty-five amino acids have been positively identified in the cation exchange column eluate¹⁰. They include eight of the protein amino acids and eleven others of known biological occurrence. The remainder appear to be unique to carbonaceous chondrites. In general terms, the Murchison amino acids comprise a mixture of C₂ through C₈ cyclic and acyclic monoamino alkanolic and alkandioic acids of apparently complete structural diversity. The carboxylic acid

Table 1 Manometric and isotopic data of amino and carboxylic acids

Run no.	Sample μ mol	Nitrogen μ mol	CO ₂ μ mol	Hydrogen μ mol	$\delta^{15}\text{N}$ (‰)	$\delta^{13}\text{C}$ (‰)	δD (‰)
Isoleucine (C ₆ H ₁₃ O ₂ N)							
1.	18.3 (solid)	8.7	113	92	3.4	-30.6	-196
2.	10.7 (solution)	4.8	70	51		-30.3	-178
3.	0.7 (solution)	0.5	4.5	4		-30.4	-195
4*	19.8 (solution)	10.0	118	126			-193
5*	Exchanged with water of $\delta\text{D}+500\%$ 8.2	4.5	53	58		-30.9	-53
6.	Amino acid blank†	<0.5	8	4		-13.9	-158
7.*	Amino acids: composite standard (solid)	8.4	75	88	7.1	-27.7	-46
8.*	Amino acid: composite standard (after desalting)	8.0	75	79	6.5	-27.7	-59
9.*	Ion exchange resin‡ 4.3 mg	8.3	151	92	20.0	-28.5	-66
10.	Amino acid fraction from Murchison meteorite	3.8	57	34	90.0	23.1	1370
11.	Carboxylic acid fraction from Murchison meteorite	6.4§	33	18	-1.3	6.7	377

$$\delta \left(\frac{R_{\text{sample}}}{R_{\text{standard}}} - 1 \right) \times 10^3 \text{ where } R = {}^{15}\text{N}/{}^{14}\text{N}, {}^{13}\text{C}/{}^{12}\text{C} \text{ and D/H respectively.}$$

Standards for nitrogen, carbon and hydrogen are air, PDB and standard mean ocean water respectively.

* Reaction vessels kept at 800 °C for two hours during extraction of H₂O. The rest of the extractions (no asterisk) were done with gentle warming of reaction vessels for short time resulting in low H₂ yields.

† A full procedural blank for preparation of the amino-acid sample.

‡ Used in isolation of amino acids.

§ Probably atmospheric contamination (see text). The blank for the extraction of gasses for isotope analyses is <0.5 μ mol for N₂ and CO₂ and about 0.5 μ mol for H₂.

Isotope data in run 10 was corrected for (1) amino acid blank (run 6) assuming the contaminant was the ion exchange resin (2) the blank for N₂, CO₂ and H₂ extraction and (3) for the easily exchangeable hydrogen (1/3 of the hydrogens). The isotope data in run 11 was corrected only for item (2). The values for the yields are uncorrected.

preparation contains the lower monocarboxylic acids in declining amounts with increasing carbon number¹¹. Both straight and branched chain isomers are present.

The procedure used for isotope analysis involved sample combustion to CO₂, H₂O and N₂ followed by reduction of the H₂O to H₂. A small apparatus giving low blank values was employed (see Fig. 1). The dissolved sample was placed in tube A and the water evaporated in an oven. A few CuO granules were added and the tube was evacuated. After about one hour the vial was sealed off and heated to 850 °C for several hours to combust the sample. Tube A was then placed in tube B which contained a small uranium chip in the side arm. The entire apparatus was thoroughly degassed, the lower part immersed in liquid N₂, and tube A cracked with the magnetically operated breaker. The released N₂ was pumped into a sample tube. The liquid N₂ bath was replaced by a Dry Ice bath and the released CO₂ was condensed in a second sample tube. Finally H₂O was frozen out in the side arm and sealed off. After heating to about 700 °C to convert H₂O to H₂, the sealed side arm was placed in a vessel similar to that of Fig. 1, cracked, and the H₂ pumped into a third sample tube.

This procedure produces 100% yields of N₂ and CO₂ within experimental error of ± 1 μ mole but considerably lower yields for H₂O. Low H₂O recovery can be attributed to the hydration of the silica in the combustion process and to the difficulty of recovering the water of hydration. Heating the reaction tube with a furnace at 800 °C for two hours during the H₂O transfer resulted in quantitative recovery.

The results of the isotope and manometric gas analyses are given in Table 1. The run numbers marked with an asterisk indicate that the H₂O extractions were done at high temperature. Runs 1, 2, 3 and 4 are the data for control experiments carried out with an isoleucine standard and demonstrate the applicability of the isotopic analytical technique for amino acids in the abundance range of the Murchison meteorite sample. Run 5 addresses the question of exchangeable hydrogens and confirms that about three of the thirteen hydrogens of isoleucine are isotopically exchangeable.

The next four runs are control experiments associated with isolation of the Murchison amino acids. Run 6 is the total procedural blank. The N₂, CO₂ and H₂ produced here are largely

due to 'bleed' from the ion-exchange resin. Comparison of runs 7 to 8 could indicate that the ion-exchange elution step contributed additional carbon, hydrogen or nitrogen when an amino acid standard similar in composition to the Murchison amino acids was tested. This would be true if the loss of hydrogen is due partially to the loss of some of the standard amino acids in the elution step. The δ values of the ion-exchange resin (run 9) indicate that as a contaminant, it would decrease the δ values of the extracted meteorite amino acids.

Run 10 shows the analyses of amino acids from the Murchison meteorite. All δ values have been corrected for appropriate blanks. The δD value has been corrected for the presumed loss of D by exchange with the isotopically lighter terrestrial water during the last solution of the amino acids. This correction is based on the assumption that each amino acid has three readily exchangeable protons (the α -ammonium protons of the zwitterion). It can be shown that these represent one-third of the total amino acid hydrogen for an amino acid mixture of the Murchison meteorite composition. In addition, it is known that some exchange of hydrogen at stable C-H bonds occurs during hot-water extraction of the amino acids from the meteorite powder¹². Although no correction for this effect was attempted, it is clear that the meteorite amino acids should have a more positive δD value than those reported here.

The yield of H₂ from the Murchison amino acid sample is low because the water transfer after combustion was not done at high temperature. This is well demonstrated by comparing the low temperature yields of runs 1 and 2 with runs 4 and 5. However, it should be noted that δD values for the isoleucine standard are the same for runs 1 and 4 despite similar temperature-related differences in H₂ yield. Thus, incomplete transfer of the water of combustion should have no bearing on the isotopic data for either the amino acids or the monocarboxylic acids. Ion-exchange column 'bleed', which may be present in the amino acid sample, could also contribute to the apparently low H/C ratio since this material yields substantially more CO₂ than H₂ (see run 9). The δ values we report for the Murchison amino acids should be considered minimum values since (1) the presence of a contaminant would lower the δ values and (2) in the case of δD , isotope exchange is only partly corrected for. It is clear from these data that the Murchison amino acids

are extraterrestrial in origin. Their high δD value is similar to those of extraterrestrial and probably interstellar components present in the HCl-HF insoluble residue of the meteorite.

The δD value for the carboxylic acids (run 11) is not as unusual as that of the amino acids but is outside the range of terrestrial values. The yield of CO_2 in this case is almost exactly what is expected based on the GC analysis done prior to isotope analysis and the experimental H/C ratio, after correction for the low H_2 yield, is also close to the expected value. The $\delta^{13}C$ is in good agreement with earlier analyses of the individual acids¹¹. The presence of N_2 in the carboxylic acid preparation is inexplicable, but the similarity of its $\delta^{15}N$ value to that of atmospheric N_2 suggests the possibility of a small air leak during the transfer of the liquid nitrogen-noncondensable-fraction in the extraction procedures.

The Murchison amino acid data provide evidence that the amino acids formed at least in part from compounds that had high δD values, a characteristic of interstellar compounds. In interstellar space D-enrichment in molecules such as HCN and CH_2O is due to the partition of D/H between these molecules and H_2 by low temperature ion-molecule reactions¹³⁻¹⁴. In contrast, the δD of meteoritic inorganic hydrogen, which occurs primarily as hydrated silicates, for example, clays and serpentines, and is the dominant source of hydrogen in meteorites, has been consistently found to be roughly -100% (ref. 7). This value is similar to that of terrestrial hydrogen and to water in a solar nebula equilibrated with cosmic hydrogen (D/H $\sim 20 \times 10^{-6}$) at temperatures near 250 K (ref. 15). The δD of methane equilibrated with cosmic hydrogen at these temperatures would also be similar to terrestrial hydrogen. Thus hydrogen-containing compounds equilibrated in the solar nebula should have terrestrial δD values. The *de novo* formation of the meteorite amino acids from primordial gases either in the solar nebula or early in the history of the solar system seems a less likely possibility on the basis of these results.

It will be of great interest to extend these analyses to additional classes of organic compounds in the Murchison meteorite, to selected individual compounds, for example, particular amino acids, and to organic compounds from other carbonaceous chondrites. Differences in isotope ratios, particularly δD , may be indicative of varying degrees of chemical processing and/or mixing with products of nebular or solar syntheses. It seems clear that further δD measurements on separated organic compounds will provide a significant approach to answering the interesting and important question of the origin of meteorite organic compounds.

The relevance of meteorite organic compounds to the origin of life on Earth might be investigated using N, H, and C isotope data from very ancient, well preserved rocks. Careful combined chemical and isotope analyses of amino acids or related organic compound in the most ancient cherts may provide some clues on these matters.

Work at the California Institute of Technology was supported by NASA grant NAG 9-46, and work at Arizona State University was supported by NASA grant NSG 7255.

Received 18 August 1986; accepted 26 January 1987.

1. Cronin, J. R., Gandy, W. E. & Pizzarello, S. in *Biogeochemistry of Amino Acids* (eds Hare, P. E., Hoering, T. C. & King, K. Jr) (Wiley, New York, 1980).
2. Kvenvolden, K. A. *et al. Nature* **228**, 923-926 (1970).
3. Smith, J. W. & Kaplan, I. R. *Science* **167**, 1367-1370 (1970).
4. Chang, S. & Mack, R. *Lunar planet. Sci.* **9**, 157-159 (1978).
5. Becker, R. H. & Epstein, S. *Geochim. cosmochim. Acta* **46**, 97-103 (1982).
6. Robert, F. & Epstein, S. *Geochim. cosmochim. Acta* **46**, 81-95 (1982).
7. Yang, J. & Epstein, S. *Geochim. cosmochim. Acta* **47**, 2199-2216 (1983).
8. Lewis, R. S., Anders, E., Wright, I. P., Norris, S. J. & Pillinger, C. T. *Nature* **305**, 767-771 (1983).
9. Kerridge, J. F. *Earth planet. Sci. Lett.* **64**, 186-200 (1983).
10. Cronin, J. R. & Pizzarello, S. *Geochim. cosmochim. Acta* **50**, 2419-2427 (1986).
11. Yuen, G., Blair, N., DesMarais, D. J. & Chang, S. *Nature* **307**, 252-254 (1984).
12. Pereira, W. E. *et al. Geochim. cosmochim. Acta* **39**, 163-172 (1975).
13. Wannier, P. G. A. *Rev. Astr. Astrophys.* **18**, 399-437 (1980).
14. Penzias, A. *Science* **208**, 663-669 (1980).
15. Geiss, J. & Reeves, H. R. *Astr. Astrophys.* **93**, 189-199 (1981).

Rubber state of ionic fluorozirconate glasses

K. Abbas*, C. Mai*, S. Etienne*, J. Perez* & G. P. Johari†

* Groupe d'Etude Metallurgie Physique et de Physique des Matériaux, LA341, Institut National des Sciences Appliquées, 69621-Villeurbanne, France

† Department of Materials Science and Engineering, McMaster University, Hamilton, Ontario L8S 4L7, Canada

Long-chain, and network polymers are elastic over a range of temperature, known as the rubbery or plateau region¹. On cooling they become rigid, or glassy, with a shear modulus of ~ 10 GPa and, on heating, if crystallization or decomposition does not intervene non-network polymers become fluid with a shear modulus approaching zero. In the rubbery zone, the real component of their shear modulus, G' , remains nearly constant and the mechanical loss factor, $\tan \phi$, reaches a minimum value¹. The occurrence of the rubbery state is exclusively a property of polymers in which topological constraints, resulting from chain entanglements, or junction points forming the network, exist². These obviously involve flexible covalent bonds at junction points and intermolecular or interchain forces at entanglements. Here, we report the first observation of rubber-like behaviour in an ionic fluorozirconate glass³, currently of great interest for use as optical waveguides and infrared windows, and examine its implications for concepts of polymer rheology. These results are significant for they suggest that ionic bonds in non-polymers can provide the same configurational restrictions as covalent bonds in a macromolecule or a polymer.

Measurements of modulus, G' , and $\tan \phi$ were made on samples of dimensions $\sim 6 \times 50 \times 1$ mm at 2 K intervals over a temperature range of 77 K-620 K at frequencies of 10^{-2} , 10^{-1} and 1 Hz with an inverted forced oscillation pendulum described earlier^{4,5}. Measurements were made on samples held in vacuum. Several different compositions of ZrF_4 -based glass were studied. Typical dynamic mechanical behaviour of 0.55 ZrF_4 ; 0.30 BaF_2 ; 0.07 NaF ; 0.05 LaF_3 ; 0.03 AlF_3 glass is given here. Results for other compositions were similar.

G' and $\tan \phi$ of the fluorozirconate glass measured at 0.01 Hz are plotted against temperature in Fig. 1. The results of measure-

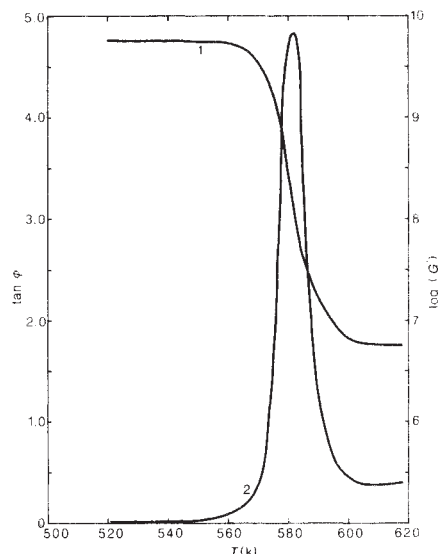


Fig. 1 The modulus, G' , (curve 1) and loss factor, $\tan \phi$ (curve 2) of a fluorozirconate glass measured at 0.01 Hz plotted against temperature. The glass transition temperature, $T_g = 580$ K, is measured by differential scanning calorimetry. G' is in Pa.

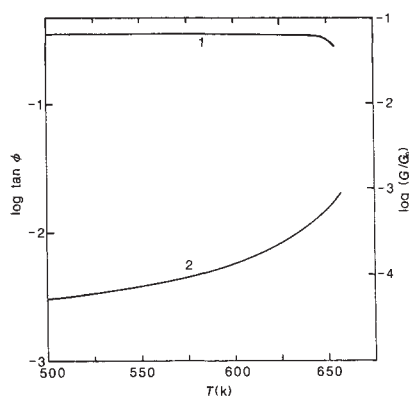


Fig. 2 The G' (curve 1) and $\tan \phi$ (curve 2) of a partially crystallized fluorozirconate glass measured at 0.01 Hz. $G_0 = 20$ GPa.

ments at 0.01 Hz, exclusively given in Fig. 1, indicate that the glass was sufficiently stable against crystallization above its T_g for the appreciably long period of time needed for the measurements at this frequency. The corresponding isochrones at higher frequencies were qualitatively similar but were shifted to higher temperatures. This shift in temperature of the inflexion point of G' and of the $\tan \phi$ peak corresponded to a thermally activated process which is characteristic of materials near T_g . G' and $\tan \phi$ remained unchanged on recycling the sample through $350 < T < 620$ K and its X-ray diffraction taken after the measurements showed no crystallinity in it. Further examination under an optical microscope indicated no surface crystallization. The calorimetric T_g of the fluorozirconate is ~ 580 K, which agreed with the temperatures of both the $\tan \phi$ peak and the inflexion point of G' isochrone in Fig. 1. Above the transition zone, G' remains constant at $T > 600$ K, a feature characteristic of the rubbery-, or plateau-, region.

One sample was partially crystallized by keeping it for 4 h at 620 K, and again measured. A plot of its G' and $\tan \phi$ in Fig. 2 shows that partial crystallization increased its G' by a factor of 10 and reduced $\tan \phi$ by a factor of 100.

Measurements of G' and $\tan \phi$ at a frequency of 1 Hz at low temperatures, given in Fig. 3, showed the persistence of localized motions in the rigid matrix of the fluorozirconate glass, as a β -, or secondary, relaxation. The temperature of the mechanical β -relaxation peak measured at 1 Hz is 273 K, nearly half its T_g . Almost all amorphous polymers⁶ measured at 1 Hz show a β -relaxation peak at about the same fraction of T_g .

In Fig. 1, the modulus, G' , of the glass is seen to decrease from ~ 6 GPa at 550 K to ~ 7 MPa at ~ 620 K. The three-decade decrease in G' is generally regarded as characteristic of a long-chain, or network, polymer^{1,2,6} and arises from topological constraints in such structures. In contrast, the modulus of inorganic network glass, for example, SiO_2 , GeO_2 , Si approaches zero at $T \geq T_g$.

Deduced from vibrational spectroscopy and diffraction results^{7,8}, the structure of a binary $\text{BaF}_2 \cdot 2\text{ZrF}_4$ glass is thought to consist of randomly arranged ZrF_6 octahedra sharing corners (or a doubly bonded F atom) and possibly edges^{8,9,10}, thus forming zigzag chains of octahedra. The nature of the Zr-F-Zr bond is uncertain but it is referred to as 'strong bond'^{9,10}. The chains are 'cross-linked' by Ba-F ionic bonds or by doubly bonded F atoms. Thus the structure resembles an ionically bonded network similar to those of metasilicates, metaphosphates, GeO_2 , for example, where four to eight membered rings exist. Such structures do not show a rubber-like state. Owing to its different composition and the presence of LaF_3 and AlF_3 , the fluorozirconate of our study may not have the same structure

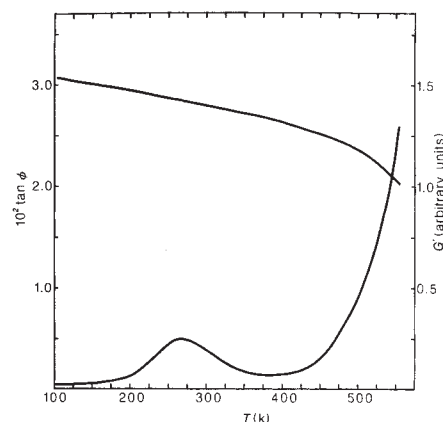


Fig. 3 The G' (curve 1) and $\tan \phi$ (curve 2) of the fluorozirconate glass measured at 1 Hz showing the β -relaxation peak. G' is in arbitrary units.

as $\text{BaF}_2 \cdot 2\text{ZrF}_4$, but its rubber-like state does suggest that the chains in it must be long enough to cause entanglement network type topological constraints.

Chain entanglements, network and transient cross-links, existing over the timescale of the measurement, are necessary for the observation of a rubber state of an inorganic or organic polymer^{1,2,6}. In the fluorozirconate structure, permanent cross-links or junction points between the assumed chains cannot exist because they would not allow long-range translational diffusion of the whole molecule necessary for nucleation and crystal growth—and we observed its partial crystallization (Fig. 2) after a period of 4 h at ~ 620 K, but transient cross-links showing longer time arrangements can exist. Thus its rubber-like behaviour must be due to the extensive entanglements of the assumed chain-like structures and/or the presence of transient cross-links in it. The concepts for both the entanglement network^{1,2} and the deGennes-Edwards' type¹¹⁻¹³ restraint of the chain (polymer molecule) in an effective tube is associated with a terminal relaxation time, which is much longer than it would be in the absence of such topological restraints. This longer time, which corresponds to the terminal zone, may control the partial crystallization of the rubbery fluorozirconate observed here.

The rheological behaviour of polymers is attributed to the configurational changes involving both the intramolecular barriers which are caused by the directionality and the fixed distance of covalent bonds and the interchain interactions, which constrain the movement of the chain within the tube. The intra- and inter-'chain' interactions in fluorozirconates are strongly ionic and therefore less directional. Yet, their rheology is strikingly similar to that of covalently bonded inorganic and organic polymers.

Received 1 September 1986; accepted 2 January 1987.

1. Ferry, J. D. *Viscoelastic Properties of Polymers* 3rd edn, (Wiley, New York, 1980).
2. Treloar, L. R. G. *The Physics of Rubber-like Elasticity*, 3rd edn (Clarendon, Oxford, 1975).
3. Poulain, M. *Nature* **293**, 279-280 (1981).
4. Cavaille, J. Y. & Etienne, S. *Mém. scient. Revue Métall.* **383-418** (1984).
5. Etienne, S., Cavaille, J. Y., Perez, J., Point, R. & Salvia, M. *Rev. scient. Instrum.* **53**, 1261-1264 (1982).
6. McCrum, N. G., Read, B. & Williams, G. *Anelastic and Dielectric Effects in Polymeric Solids* (Academic, New York, 1967).
7. Almeida, R. M. & MacKenzie, J. D. *J. chem. Phys.* **74**, 5954-5961 (1981); *J. Non-Cryst. Solids* **51**, 187-199 (1982).
8. Lucas, J., Angell, C. A. & Tamaddon, S. *Bull. mater. Res.* **19**, 945-951 (1981).
9. Coupé, R. R., Louër, D., Lucas, J. & Leonard, A. *J. Am. Ceram. Soc.* **66**, 523-529 (1983).
10. Kawamoto, Y. & Horisaka, T. *J. Non-Cryst. Solids* **56**, 39-44 (1983).
11. deGennes, P. G. *J. chem. Phys.* **55**, 572-579 (1971); *Physics Today* **36**, 33-39 (1983).
12. Doi, M. & Edwards, S. F. *JCS Faraday Trans. II* **74**, 1789-1801, 1802-1817, 1818-1832 (1978); **75**, 38-54 (1979).
13. Graessley, W. W. *Adv. Polym. Sci.* **47**, 67-117 (1982).

Concentrations and dry deposition velocities of nitric acid

G. J. Dollard, D. H. F. Atkins, T. J. Davies & C. Healy

Environmental and Medical Sciences Division,
Harwell Laboratory, Oxfordshire OX11 0RA, UK

During recent years much effort has concentrated on the physical and chemical behaviour of the oxides of nitrogen and sulphur in the atmosphere in order to aid understanding of the formation, transport and deposition of acidity on to potentially sensitive ecosystems. Nitric acid is an important product of atmospheric nitrogen chemistry but relatively little information is available on which to judge its fate in the atmosphere. Here, we present some preliminary data on the gas phase concentrations of nitric acid for a rural site in central southern England. Measurements of the dry deposition flux of this acid are also given. The concentration of nitric acid was found to vary both diurnally and seasonally. From the dry deposition measurements it was found that there was little or no surface resistance to deposition and derived deposition velocities were considerably higher than those often quoted for other pollutant gases. It is concluded that the dry deposition flux of nitric acid constitutes an important input of acidity to ground surfaces.

During recent years much work has been directed towards elucidating the mechanisms causing and controlling atmospheric acidity. The chemistry of acid formation in the troposphere is complex involving both primary pollutants (NO , SO_2) and secondary pollutants (NO_2 , HNO_3 and H_2SO_4) in an intricate matrix of reactions. These reactions involve other trace species, may occur in gas or liquid phases and are profoundly influenced by solar radiation^{1,2}.

The overall effect of these reactions is the maintenance of a generally oxidizing environment in which nitrogen and sulphur oxides present in the atmosphere form nitric and sulphuric acids respectively. These strong acids account for most of the acidity present in rain water. The deposition to the ground of these species in rain may be measured using established rain-water collecting techniques³. This deposition is augmented by the direct, dry deposition of both primary and secondary pollutants. Relatively little is known either of the concentration of nitric acid in the gas phase or the rate of its removal from the atmosphere by dry deposition.

In all the work reported below the concentrations of HNO_3 were determined using an open-face nylon filter technique⁴. This method relies upon the quantitative removal of the gaseous acid by its reaction with the amine groups of the adipic acid/hexamethylene diamine copolymer of the filter (nylon 66, 0.45 μm pore size) after removal of particulate species by a 1 μm pore size Teflon pre-filter. Nylon filters are known to be highly efficient at removing nitric acid from ambient air^{4,5}. Variable background nitrate levels on the nylon filters were reduced to acceptable levels (2 $\mu\text{g NO}_3^-$ per filter) by the routine pre-treatment of the filters by repeated extraction with deionized water in a Soxhlet apparatus.

Nitrate collected on the nylon filters was extracted with 0.005 M NaOH under reflux and determined on a continuous flow auto-analyser system⁶.

It has been found that under high particulate loadings nitric acid may be lost from the prefilter and collected on the nylon filter⁷. This is due to the dissociation of solid ammonium nitrate, a process affected by both temperature and humidity^{8,9}. However, at the loadings experienced under field sampling conditions such losses are not considered to constitute an important source of error¹⁰.

Two series of experiments are described below; a year-long study of concentrations of gaseous HNO_3 , which was carried

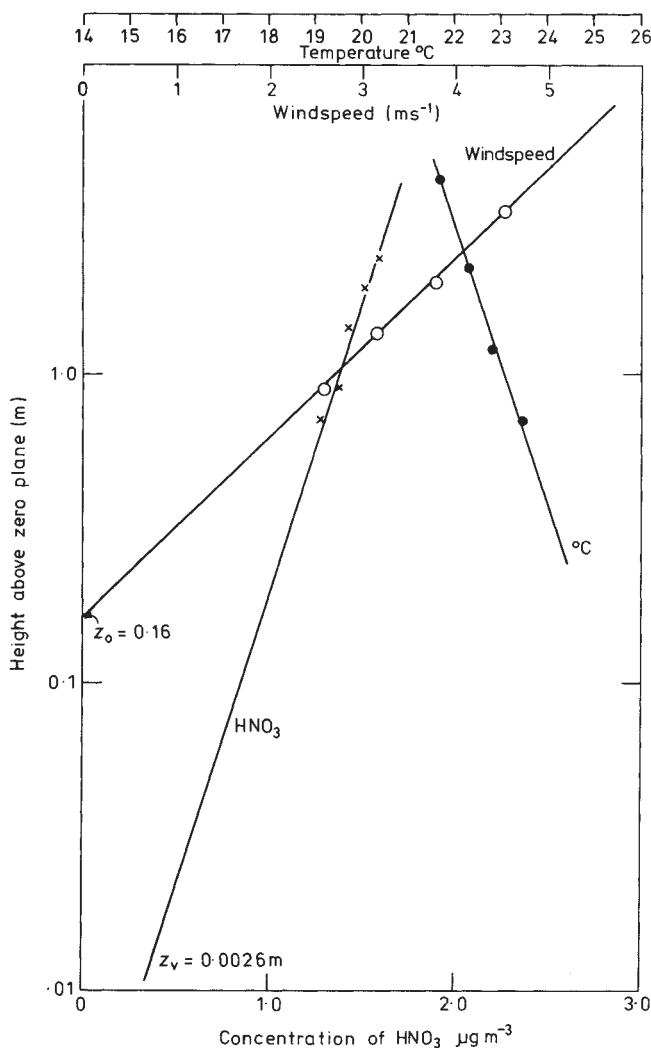


Fig. 1 Examples of profiles used to derive deposition velocities in the micrometeorological study. z_0 , Roughness length; z_v , analogue of z_0 for acid vapour.

out within the grounds of The Harwell Laboratory and measurements of dry deposition rates of these gases to a cereal crop grown at a commercial farm a few miles from Harwell.

For the year-long study of ambient concentrations filter packs (47-mm diameter) were mounted 1 m above a grass surface and operated at a sampling rate of 15 l min⁻¹ for periods of 8 to 16 hours. Sampling was carried out randomly throughout the week and included both daytime (sunrise to sunset) and night-time periods.

The results of this study are summarized in Table 1. Both seasonal and diurnal variations are evident in the HNO_3 data. This behaviour may be attributed to several factors. Firstly the photochemical production of HNO_3 ^{1,2} will vary with the intensity of solar radiation; secondly, temperature and humidity may alter the dissociation equilibrium of ammonium nitrate^{8,9}, and thirdly, loss of HNO_3 from the nocturnal boundary layer due to deposition must occur. The third effect is considered to be the most important in defining the day/night ratios. When the nocturnal inversion is established there is little turbulent transport to the surface layers. As a consequence, material lost to the ground is not replaced, leading to a fall in concentration as has been noted for ozone¹¹ and PAN¹² (peroxy acetyl nitrate). Seasonal changes in atmospheric ammonia concentrations may also influence the nitric acid-ammonia-ammonium nitrate equilibrium. Ammonia is important in this respect as the resi-

Table 1 Summary of HNO₃ measurements made at the Harwell Laboratory

Periods*	Number of determinations	Mean HNO ₃ μg m ⁻³
Year	76	1.65 (0.2–12.5)†
Summer	44	2.10
Winter	32	1.05
Summer day	23	3.20
Summer night	21	0.91
Winter day	24	1.20
Winter night	8	0.66

* Daytime defined as sunrise to sunset. Summertime April–September inclusive.

† Range of measured concentrations.

dence time of nitrate in the atmosphere is dependent to a large extent on whether the nitrate is in the gaseous or particulate form.

The measured concentrations of nitric acid are similar to the few reported for the UK^{8,13} and comparable to values reported for rural locations in the USA^{10,14}, northern Europe¹⁵ and central Europe¹⁶.

Dry deposition velocities were obtained using a micrometeorological technique. Windspeed and temperature were recorded at five heights, above a mature wheat crop, using cup anemometers and platinum resistance thermometers. Concentrations of the gaseous acids were measured using 90 mm filter packs positioned at 5 heights above the surface. Measurements were made over periods of 30 to 120 minutes at a flowrate of 40 l min⁻¹. During July and August 1984 20 sets of measurements were completed. Typical profiles are shown in Fig. 1.

The deposition flux densities, F , were derived from the gradient of concentration with height, $d\chi/dz$, assuming that the gases were transported by turbulence similarly to heat:

$$F = k_H(z) d\chi/dz$$

where the eddy diffusivity, for heat at height z , $k_H(z)$ was obtained from the profiles of windspeed and height with appropriate stability corrections following Garland¹⁷.

In Table 2 the results are expressed as the deposition velocity

$$V_g(z) = F/\chi(z)$$

choosing a value of 1 m above d , the zero plane, for the reference height, z .

Some indication of the importance of the mechanisms controlling the deposition rate can be achieved by partitioning the total resistance, $r_t = V_g(z)^{-1}$ between aerodynamic and surface terms¹⁷. The aerodynamic resistance terms $r_a(z)$ and r_b represent the control exerted by the rate of transport through the atmosphere from height z , to the surface. These terms may be evaluated from the micrometeorological measurements¹⁷. Any remaining resistance is attributed to the surface itself; very reactive gases would be expected to have little surface resistance.

The deposition velocities derived for HNO₃ are in the range 0.05 to 0.26 m s⁻¹ and are substantially larger than the range of values (0.005–0.01 m s⁻¹) commonly quoted for SO₂ and other reactive gases¹⁸.

The scatter in the deposition velocity data is difficult to attribute to any meteorological factors. However, it is possible that on occasions there could be a conversion of HNO₃ into ammonium nitrate aerosol near to the crop surface. This could be a consequence of the release of ammonia from the soil–crop system following biological activity or emission from fertilizer residues. Such a depletion of HNO₃ near to the surface would generate apparently stronger gradients. From the measurements made at the time it is not possible to assess these possibilities. However, experiments are in hand that will clarify the situation.

Considering, on average the data in Table 2, it can be said

Table 2 Summary of deposition measurements

Run	Time GMT	z_0 m	u_* m s ⁻¹	r_a s m ⁻¹	HNO ₃	
					Flux μg m ⁻² s ⁻¹	V_g m s ⁻¹
2	1430	0.15	0.54	7.90	1.47	0.26
3	1123	0.15	0.53	8.10	1.03	0.11
6	1231	0.10	0.29	13.0	0.78	0.11
7	1347	0.10	0.37	13.20	0.80	0.075
9	1213	0.12	0.62	7.70	0.33	0.164
10	1320	0.13	0.73	6.26	0.44	0.191
11	1432	0.12	0.61	8.06	0.47	0.181
15	1444	0.19	0.67	5.86	0.07	0.049
16	1600	0.22	0.71	5.68	0.04	0.081
17	1723	0.14	0.70	6.67	0.05	0.133

z_0 , roughness length; u_* , friction velocity.

that the r_b and r_a terms are small and the results are consistent with HNO₃ having little surface resistance to deposition.

There are no published values for the deposition velocities of nitric acid to tall vegetation measured in the UK but values have been published for the deposition of nitric acid to a grass surface in the USA^{10,14}, ranging from 0.01 to 0.047 m s⁻¹. The higher range indicated in the present study is attributable to the enhanced turbulent diffusion over the aerodynamically rougher wheat crop. Recent measurements of nitric acid deposition to a forest canopy¹⁹ indicate deposition velocities equivalent to the high values obtained in the present study. Again there was little evidence for any surface resistance to nitric acid deposition. The fact that the deposition velocities were similar over two different surfaces reflects similar conditions of turbulence over each surface (values of the friction velocity were similar) and is indicative of the importance of aerodynamic factors in the deposition of reactive species.

Analysis of the Teflon prefilters for particulate species was complicated by the variable background levels of nitrate on the filters (which could not be pre-treated owing to their tendency to crease and curl once removed from the backing paper). Particulate nitrate was variable and ranged from about 32% of the total nitrate present to about 75%, due mostly to a fall in gas phase concentrations of HNO₃.

The fraction of nitric acid in the gas phase (up to ~70%) is higher than that previously reported for wintertime in the UK⁸ probably reflecting the effect of higher ambient temperatures upon the dissociation of ammonium nitrate. Simultaneous measurement of ammonia during the deposition experiments allowed calculation of the equilibrium constants for the dissociation; these were found to be in relatively good agreement with theoretical predictions for ammonia and nitric acid in equilibrium with solid ammonium nitrate at high ambient temperatures and low humidity^{8,9}.

The dry deposition of HNO₃ is important as it facilitates the direct input of acidity to ground surfaces. Dry deposited SO₂

Table 3 Dry deposition of HNO₃, and the strong acid precursors NO₂ and SO₂ together with the wet deposited H⁺ for a hypothetical cereal crop in central southern England (Harwell)

	Dry deposition			Wet deposition
	HNO ₃	NO ₂	SO ₂	H ⁺ §
Concentrations*	1.65	28†	20‡	40
Deposition velocity m s ⁻¹	0.135	0.0025§	0.0085¶	
Flux meq m ⁻² hr ⁻¹	0.012	0.006	0.019	0.003

* Units of concentration are μg m⁻³ of air for the gases and vapours, listed under dry deposition and mg l⁻¹ of rain water for the ions listed under wet deposition.

† DHF Atkins unpublished data; ‡ ref. 13; § ref. 3; ¶ ref. 17.

is generally oxidized to H_2SO_4 with each molecule having the potential to contribute two hydrogen ions. Similarly each molecule of NO_2 deposited can yield one hydrogen ion.

The relative proportions of these components contributing to acid deposition to a hypothetical cereal crop in southern England are shown in Table 3. Dry deposited SO_2 comprises the greatest fraction of the deposited acidity followed by the contribution from HNO_3 , and the magnitudes of all the dry deposited species are individually larger than the 'acid rain' component.

These simple estimates serve to underline the importance of dry deposition mechanisms in the direct transfer of strong acids to vegetation surface structures. The aerodynamic resistance to transfer to forests and cities is considerably smaller than that for cereal crops. If similar (or even higher) concentrations of HNO_3 were maintained the deposition of these species would be correspondingly increased.

This work was supported by the Department of the Environment. We are grateful to J. A. Garland for advice and discussions.

Received 11 June 1986; accepted 9 February 1987.

1. Cox, R. A. *J. Photochem.* **25**, 43-48 (1984).
2. Cox, R. A. & Penkett, S. A. in *Acid Deposition* (eds Beilke, S. & Elshor, A. J.) 58-83 (CEC, Brussels, 1983).
3. *UK Acid Rain Review Group Report* (Warren Spring Laboratory, Stevenage, 1983).
4. Goldan, P. D. *et al. Atmos. Environ.* **17**, 1355-1364 (1983).
5. Grosjean, D. *Anal. Lett.* **15**, (A9) 785-796 (1982).
6. Eggleton, A. E. J. & Atkins, D. H. F. *Results of the Teeside investigation UKAEA Report AERE-R 6983* (HMSO, London, 1972).
7. Appel, B. R., Tokiwa, T. & Haik, M. *Atmos. Environ.* **15**, 283-289 (1981).
8. Harrison, R. M. & Pio, C. A. *Tellus* **35B**, 155-159 (1983).
9. Stelson, A. W., Frielander, S. K. & Seinfeld, J. H. *Atmos. Environ.* **13**, 369-371 (1979).
10. Huebert, B. J. & Robert, C. H. *J. geophys. Res.* **90**, (D1) 2085-2090 (1985).
11. Garland, J. A. & Derwent, R. G. *Q. J. R. met. Soc.* **105**, 169-183 (1979).
12. Garland, J. A. & Penkett, S. A. *Atmos. Environ.* **10**, 1127-1131 (1976).
13. Cox, R. A., Derwent, R. G. & Sandalls, F. J. *Some air pollution measurements made at Harwell, Oxfordshire during 1973-1975*. UKAEA Report AERE-R 8329 (HMSO, London, 1976).
14. Huebert, B. J. in *Precipitation Scavenging, dry deposition and resuspension* Vol. 2 (Coordinators Pruppacher, H.R. Semonin, R. G. & Slinn, W. G. N.) 785-793 (Elsevier, New York, 1983).
15. Grennfelt, P. *Atmos. Environ.* **14**, 311-316 (1980).
16. Meixner, F. X., Muller, K. P., Aheimer, G. & Hofken, K. D. *Proc. COST action 611 Meet: 'Pollutant cycles and transport: modelling and field experiments'* (eds de Leuw, F. A. A. M. & van Egmond, N. D.) 103-119 (RIVM Bilthoven, Netherlands, 1985).
17. Garland, J. A. *Proc. R. Soc. A354*, 245-268 (1977).
18. Schmel, G. A. *Atmos. Environ.* **14**, 983-1011 (1980).
19. Meixner, F. X. in *Proc. 2nd Meet. on Atmosphere Biosphere Interactions Mainz, 1986* (in the press).

The dependence of hurricane intensity on climate

Kerry A. Emanuel

Center for Meteorology and Physical Oceanography,
Massachusetts Institute of Technology, Cambridge,
Massachusetts 02139, USA

Tropical cyclones rank with earthquakes as the major geophysical causes of loss of life and property¹. It is therefore of practical as well as scientific interest to estimate the changes in tropical cyclone frequency and intensity that might result from short-term man-induced alterations of the climate². In this spirit we use a simple Carnot cycle model to estimate the maximum intensity of tropical cyclones under the somewhat warmer conditions expected to result from increased atmospheric CO_2 content. Estimates based on August mean conditions over the tropical oceans predicted by a general circulation model with twice the present CO_2 content yield a 40-50% increase in the destructive potential of hurricanes.

We shall first review a simple model³ of the mature tropical cyclone that is capable of predicting the maximum intensity that can be achieved by such storms as a function of environmental conditions. We then apply this model to the estimation of maximum cyclone intensity under a range of climates, and to a

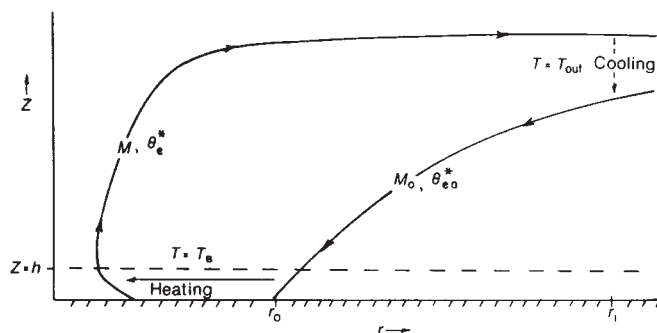


Fig. 1 Carnot cycle of the mature tropical cyclone³. Air begins with absolute angular momentum per unit mass M_0 and moist entropy θ_{e0}^* at a radius r_0 , and flows inward at constant temperature T_e within a thin boundary layer, where it loses angular momentum and gains moist entropy from the sea surface. It then ascends and flows outward to large radii, preserving its angular momentum (M) and moist entropy (θ_e^*). Eventually, at large radii, the air loses moist entropy by radiative cooling to space at a mean temperature T_{out} and acquires angular momentum by interaction with the environment.

specific situation predicted by a general circulation model simulation with increased atmospheric CO_2 .

Tropical cyclones make textbook examples of the operation of a Carnot engine, with the provision that the heat input is largely in the form of the latent heat of vaporization. Figure 1 illustrates the energy cycle³. At some radius r_0 (typically 100-500 km) surface air begins to flow inwards towards the cyclone centre within a frictional boundary layer whose depth is ~ 1 or 2 km. During its inward trek the air maintains a nearly constant temperature that is very close to the sea surface temperature but acquires water vapour from the ocean, which supplies the latent heat of vaporization. The rate at which it acquires latent heat is a monotonic function of the near-surface wind speed. As the air flows toward lower pressure, heat is added due to isothermal expansion as well. It is also during this inflow that air suffers the greatest frictional dissipation. At a much smaller radius (typically from 5 to 100 km) the air abruptly turns upward and ascends through the deep cumulo-nimbus clouds that constitute the eyewall. During this ascent the total heat content (sensible plus latent) is approximately conserved, though in the process there are large conversions of latent to sensible heat. There is also comparatively little frictional loss of energy in this branch. Finally, the air flows outward at the top of the storm and eventually loses the heat gained from the ocean by longwave radiation to space, and through friction re-acquires the lost angular momentum. This usually happens at some radius r_1 which is much larger than r_0 . Calculations are rather insensitive to the exact value of r_1 . The whole process is considered to take place in an atmosphere which is neutral to buoyant and centrifugal convection along angular momentum surfaces.

In the steady state, the mechanical energy available from this thermodynamic cycle balances frictional dissipation. As previously mentioned, the bulk of the latter occurs in the boundary layer inflow where air crosses surfaces of constant absolute angular momentum. (A much smaller amount occurs at large radii in the outflow.) The balance may be expressed symbolically

$$W_{BL} = \epsilon \Delta Q \quad (1)$$

where W_{BL} is the work done against friction in the boundary layer, ΔQ is the total heat gain from the sea surface and ϵ is the thermodynamic efficiency, which is proportional to the difference between the temperature at which the heat is added (the sea surface temperature) and a weighted mean temperature at which it is lost in the upper atmosphere (see ref. 3 for an exact definition). The efficiency ϵ is largest in the tropics, where the depth of the moist convecting layer is large, and smaller at higher latitudes, approaching zero in many places.

Knowledge of the frictional loss in the boundary layer can be used to calculate the total drop in pressure in towards the eye of the storm through the use of Bernoulli's energy equation, which may be written

$$\frac{1}{2}\Delta v^2 + \int \alpha dp + W_{BL} = 0 \quad (2)$$

where $\frac{1}{2}\Delta v^2$ is the total change in kinetic energy per unit mass following the air flowing inward in the boundary layer, α is the volume per unit mass and p is the pressure. Since v is approximately zero at the beginning of the inflow (at r_0) and is exactly zero at $r=0$, Bernoulli's equation integrated all the way into the centre of the storm may be expressed

$$\int_{r_0}^0 \alpha dp = -\epsilon \Delta Q, \quad (3)$$

where we have used equation (1) for W_{BL} . Using the ideal gas law to express α as a function of p and T (and water vapour content, which is variable), the author³ showed that the integral in equation (3) can be evaluated exactly. Specifically, both α and ΔQ can be regarded as functions of temperature, pressure and relative humidity (RH), so that equation (3) has the functional form

$$\Delta G(p, T, RH) = \epsilon \Delta F(p, T, RH) \quad (4)$$

(The exact form is given by equation (21) of ref. 3.) To a very good approximation, the temperature is a constant equal to the sea surface temperature. Moreover, the relative humidity cannot exceed unity. This permits an evaluation, using equation (4), of the maximum possible pressure drop in towards the eye as a function of the sea surface temperature, the ambient relative humidity, and the thermodynamic efficiency ϵ , the last of which can be calculated from a knowledge of the ambient temperature structure of the atmosphere.

Figure 2 shows the results of a calculation of this kind using September mean sea surface temperatures, relative humidities, and temperature profiles used in calculating ϵ . Also shown are the measured central surface pressures of several of the most intense tropical cyclones recorded. As storms in these regions tend to move towards the north-west, they often reach maximum intensity on the downstream side of the region of maximum potential intensity. Apparently, the dynamics of tropical storms sometimes permit them to achieve the maximum intensity that is energetically possible. The central pressure (or more precisely the difference between the ambient and central pressures) is a good measure of storm intensity and is also closely related to the maximum wind speeds in tropical cyclones⁴.

The pattern of maximum intensity in Fig. 2 strongly resembles the September mean sea surface temperature distribution, although the relationship is highly nonlinear. This is because the thermodynamic efficiency, ϵ , is itself mostly a function of sea surface temperature since the depth of the moist convecting layer is large over relatively warm water and small over colder

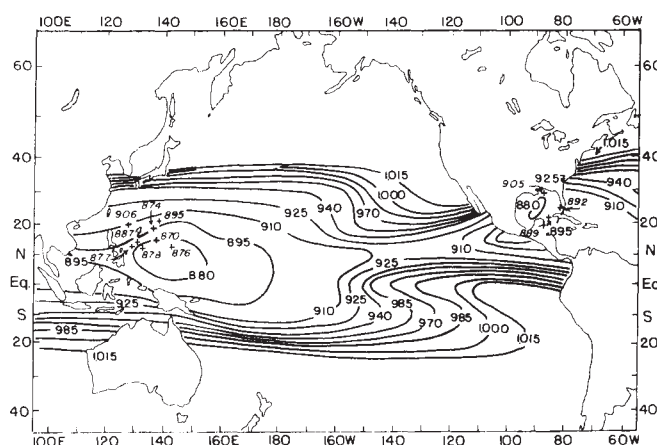


Fig. 2 Minimum sustainable central pressures (mb) under September mean climatological conditions, assuming an ambient pressure of 1,015 mb (ref. 3). Crosses mark the positions and central pressures of some of the most intense tropical cyclones on record¹.

water. The prediction of maximum cyclone intensity is therefore crucially dependent on estimates of sea water temperature. We next argue that changes in cyclone intensity due to climatic changes in ϵ and ambient relative humidity will be small.

Fractional changes in ϵ result from changes in the air temperature near the sea surface (T_B) and changes in the mean temperature of the ambient atmosphere at the levels where air flows out of the top of the storms ($\bar{T}_{out}$). The latter is typically the temperature of the lower stratosphere when ϵ is large. From the definition of ϵ

$$\frac{d\epsilon}{\epsilon} = \frac{1}{T_B - \bar{T}_{out}} \left[\frac{\bar{T}_{out}}{T_B} dT_B - d\bar{T}_{out} \right] \quad (5)$$

We are principally concerned about regions of the Earth where tropical cyclones are particularly frequent and intense; these are characterized by relatively large values of $T_B - \bar{T}_{out}$. Since this quantity has a typical magnitude of $\sim 100^\circ\text{C}$, the percentage change in ϵ will be nearly equal to the quantity in brackets in equation (5). As an upper bound on this change, consider a situation in which $\bar{T}_{out}$ changes but T_B remains fixed. It follows that a change of 4°C in $\bar{T}_{out}$ would result in only about a 4% change in ϵ . In fact, dT_B and $d\bar{T}_{out}$ are positively correlated in virtually all climate simulations performed to date. In the particular simulation discussed later in this letter, dT_B and $d\bar{T}_{out}$ are about equal. This makes fractional changes in ϵ , according to equation (5), more of the order of 1%, except where ϵ is initially small. We conclude that changes in the thermodynamic efficiency resulting from modest climate change would be insignificant.

The Carnot cycle model prediction of maximum intensity is quite sensitive to the relative humidity of near-surface air. The predictions of this quantity in large-scale numerical models is very sensitive to physical assumptions about evaporation from the sea, cumulus convection, radiation, and so on. In fact, most models do not do very well at predicting relative humidity. Nature does not appear to be as sensitive, however, since the relative humidity over the ocean is remarkably constant over a wide range of sea surface temperatures, cloud cover and precipitation. It varies from nearly 75% over colder ocean waters in middle latitudes to about 80% in the main tropical cyclone-producing regions. In many of the simpler climate models a fixed relative humidity is assumed. In view of these observations and lack of confidence in model predictions, we assume that the relative humidity over the tropical ocean remains fixed at the current values. This assumption also ties the temperature at each level in the troposphere to sea surface temperature.

With these assumptions, changes in maximum tropical cyc-

Table 1 Minimum sustainable central pressure and maximum wind speed as a function of sea surface temperature with $\epsilon = 0.33$ and an ambient relative humidity of 78%. (Ambient pressure = 1,015 mbar)

T ($^\circ\text{C}$)	P_c (mbar)	V_{max} (m s^{-1})
27	911	72
28	902	75
29	891	79
30	879	83
31	865	88
32	849	93
33	829	99
34	805	106

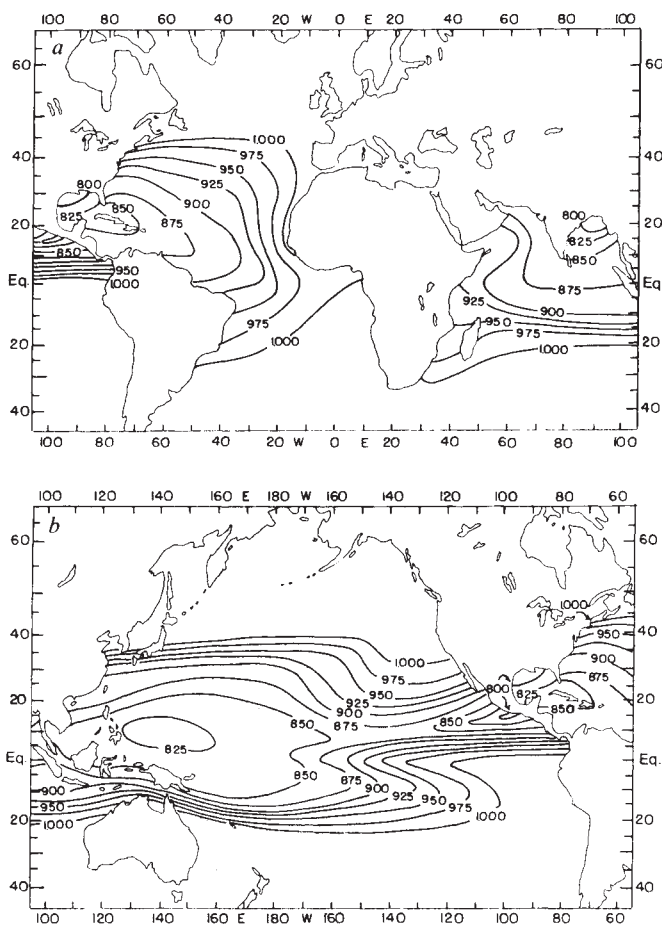


Fig. 3 Minimum sustainable central pressures (mb) as in Fig. 2, but using sea-surface temperature increases predicted by a general circulation model^{5,6} averaged over five Augusts in simulations with twice the present CO₂ content. *a*, Shows the Atlantic and Indian Oceans; *b*, the Pacific.

lone intensity are strictly related to changes in sea surface temperature. The Carnot cycle is particularly sensitive to sea surface temperature as the latent heat content of air at fixed relative humidity is a strongly exponential function of temperature, approximately doubling for each 10 °C increment of temperature above 0 °C. Table 1 shows the minimum sustainable central pressures of tropical cyclones as a function of sea surface temperature for a value of ϵ characteristic of the Caribbean and western Pacific tropical cyclone regions, assuming a surface relative humidity of 78% and an ambient surface pressure of 1,015 mb. Relatively small changes in sea surface temperature are associated with large intensity changes, with an increase of 3 °C leading to a 30–40% increase in the maximum pressure drop. Table 1 also shows estimates of the maximum wind speed in tropical cyclones based on empirical relations between wind and pressure developed by Holland⁴, taking his parameter 'B' to be 1.5. (The maximum wind is strictly bounded by the maximum pressure drop since the latter serves as the potential energy for inflowing air.) The wind increases as the square root of the pressure drop, yielding a 15–20% increase for an increase of 3 °C in sea surface temperature. The destructive potential of the wind, it should be added, varies as the square of the wind speed and thus as the total pressure drop.

To obtain some idea of the global distribution of the changes in maximum cyclone intensity that might result from predicted increases in atmospheric CO₂ we rely on predicted changes in sea surface temperature produced by the Goddard Institute for Space Studies General Circulation Model II described by

Hansen *et al.*⁵ with modifications to account for ocean mixed layer heat capacity and variable ice cover as described in ref. 6. Broadly, the model solves the conservation equations for energy, momentum, mass, and water and the equation of state on a coarse grid with a horizontal resolution of 8° latitude and 10° longitude, and with 9 layers in the vertical. The radiative scheme accounts for the radiatively significant atmospheric gases, aerosols and cloud particles, and cloud cover and height are predicted. Ground hydrology and surface albedo depend on local vegetation and snow and ice cover, which are predicted. The diurnal and seasonal cycles are included in the model. Ocean temperatures and ice cover are computed based on energy exchange between the atmosphere and an ocean mixed layer with specified heat capacity and advective heat transport. The last two vary with season at each grid point.

Figure 3 shows the minimum sustainable central pressures of tropical cyclones based on sea surface temperature changes predicted by the general circulation model averaged over five Augusts of a simulation with twice the present atmospheric CO₂ content. These changes, which in the tropics range between 2.3 °C and 4.8 °C, have been added to the current August climatological sea surface temperatures, while ϵ and surface relative humidity remain unchanged from the present climate. Minimum pressures are substantially lower in the tropics, especially in partially enclosed basins such as the Gulf of Mexico and the Bay of Bengal where minimum sustainable pressures are below 800 mbar. Were these sea surface temperature increases realized, the maximum destructive potential of tropical cyclones would be substantially increased, in some places by as much as 60%.

This analysis pertains only to the maximum sustainable pressure drop in tropical cyclones and has no direct implications for either the average intensity of cyclones or their frequency of occurrence. While one intuitively expects that average intensity increases with maximum intensity, the question of frequency is in fact poorly understood. Tropical cyclones do not arise spontaneously even under favourable conditions but appear to originate in disturbances whose origins derive from separate dynamical mechanisms, and will only do so under certain environmental conditions⁷. There is no obvious reason, however, to suppose that frequencies would be substantially diminished in a climate with doubled CO₂.

Several caveats should be mentioned regarding this analysis. In the present climate, only a small fraction of the total number of cyclones which form reach the maximum intensity given by this analysis, probably because of the relatively long spin-up time for storms of this kind, and the tendency for strong cyclonic circulations to induce upwelling of cold water. To these caveats we must add the relatively large uncertainties associated with climate simulations. In particular, the coupling of ocean dynamics to atmospheric circulation is presently quite crude. While better estimates of tropical cyclone intensities must await more sophisticated modelling efforts, the present analysis suggests that predicted climate changes associated with increased atmospheric CO₂ will lead to substantially enhanced tropical cyclone intensity.

I thank James Hansen, Gary Russell and Reto Ruedy of the NASA Goddard Space Flight Center Institute for Space Studies, and Peter Stone of MIT. The work was completed with the assistance of NSF Grant ATM-8513871.

Received 27 October; accepted 31 December 1986.

1. Anthes, A. R. *Tropical Cyclones: Their Evolution, Structure and Effects* (American Meteorological Society, Boston, 1982).
2. National Research Council (eds) *Changing Climate*, (National Academy Press, Washington, 1983).
3. Emanuel, K. A. *J. Atmos. Sci.* **43**, 585–604 (1986).
4. Holland, G. J. *Mon. Weath. Rev.* **108**, 1212–1218 (1980).
5. Hansen, J. *et al. Mon. Weath. Rev.* **111**, 609–662 (1983).
6. Hansen, J. *et al. in Climate Processes and Climate Sensitivity* (eds Hansen, J. E. & Takahashi, T.) 368 (American Geophysical Union, Washington, 1984).
7. Gray, W. M. in *Intense Atmospheric Vortices* (eds Bengtsson, L. & Lighthill, J.) 3–20 (Springer, New York, 1982).

Earth's precession cycle and Quaternary climatic change in tropical Africa

Edward M. Pokras* & Alan C. Mix†

* Lamont-Doherty Geological Observatory, Palisades, New York 10964, USA and Marine Geoscience Unit, Geological Department, University of Cape Town, Rondebosch 7700, South Africa

† College of Oceanography, Oregon State University, Corvallis, Oregon 97331, USA

High abundances of the freshwater diatom *Melosira* spp., reflecting eolian input from dry lake beds in tropical Africa, occur in deep-sea sediments of the equatorial Atlantic at discrete times during the late Quaternary^{1,2}. We report here an approximately 260 kyr record of *Melosira* accumulation in an equatorial Atlantic sediment core, and demonstrate significant influence of the Earth's precessional cycle on the climate of tropical Africa.

Eolian transport of dust from tropical Africa to the Atlantic occurs mainly in two seasonal plumes; one centred between 15° and 25°N in Northern Hemisphere summer, and one centred south of 10°N in Northern Hemisphere winter^{3,4}. To examine variations in eolian transport from tropical Africa, we have determined the abundances and accumulation rates of *Melosira* spp. in core V30-40 (0°12'S, 23°09'W, 3,706 m depth, average sedimentation rate 3 cm kyr⁻¹). At present this site receives dust from tropical Africa in Northern Hemisphere winter⁴. Our previous work¹ shows that transport of *Melosira* to the equatorial Atlantic is predominantly eolian, and that increased input of *Melosira* to the deep sea indicates increased aridity in tropical Africa.

The major strength of studying continental climates in wind-blown sediments from the deep sea is that long, continuous time series linked directly to oxygen-isotope stratigraphy can be obtained. A weakness is that the source areas of dust cannot be defined precisely, at least with data from one core. Thus, for our purposes we define tropical Africa as any area within 15° of the Equator. Based on spatial patterns of *Melosira* abundance in core-top and glacial-maximum sediments¹, it appears that the arid source areas for eolian sediments were generally in the Northern Hemisphere.

Figure 1 shows $\delta^{18}\text{O}$ stratigraphy, per-gram abundances and accumulation rates of *Melosira* spp. in core V30-40 through the last ~260 kyr. Time control is based on the correlation of the $\delta^{18}\text{O}$ record to the SPECMAP timescale (age data for V30-40 in ref. 5). *Melosira* accumulation rates are calculated using the per-gram abundances, sedimentation rates (interpolated linearly from the time-control data) and bulk densities (calculated from %CaCO₃)⁶. Differences between the time series of abundances (Fig. 1b) and accumulation rates (Fig. 1c) are small. *Melosira* maxima occur 2-4 kyr after isotopic minima in interglacial stages 1, 5, and 7, and early in glacial stage 6. Several lesser peaks occur at other times.

Based on shorter time series, we speculated previously that the *Melosira* record might contain an approximately 23 kyr period cyclic response related to the Earth's precessional period¹. With the longer time series presented here, we rigorously test the hypothesis that variations in insolation related to orbital changes affect climates in tropical Africa⁷. Results from spectral analysis⁸ (Fig. 2a) demonstrate a significant concentration of variance in *Melosira* accumulation rates at periods approximately equal to the ~23 kyr precessional period, as well as the second through fifth precessional harmonics ($p^2 = 11.5$, $p^3 = 7.3$, $p^4 = 5.5$ and $p^5 = 4.4$ kyr periods). Cross-spectral analysis⁸ (Fig. 2b) reveals that these periods in the *Melosira* time series are highly coherent (coherency = 0.86 at frequency 1/23 kyr) with

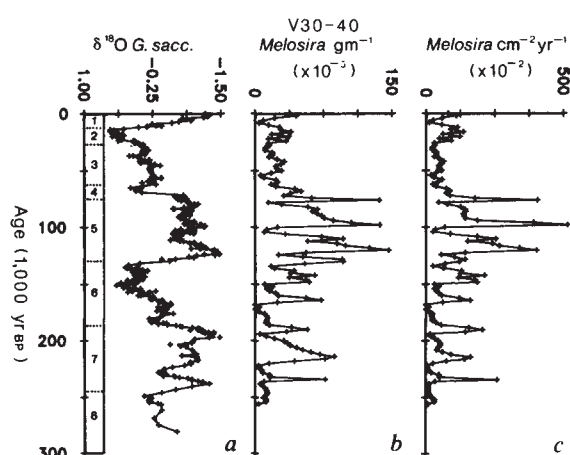


Fig. 1 a, $\delta^{18}\text{O}$ of *Globigerinoides sacculifer* (with sac. 415-500 μm), expressed in ‰ against PDB. b, Time series of *Melosira* abundances per gram ($\times 10^{-5}$) in core V30-40 through the last 260 kyr. c, Accumulation rates of *Melosira* (number $\text{cm}^{-2} \text{yr}^{-1} \times 10^{-2}$) in core V30-40.

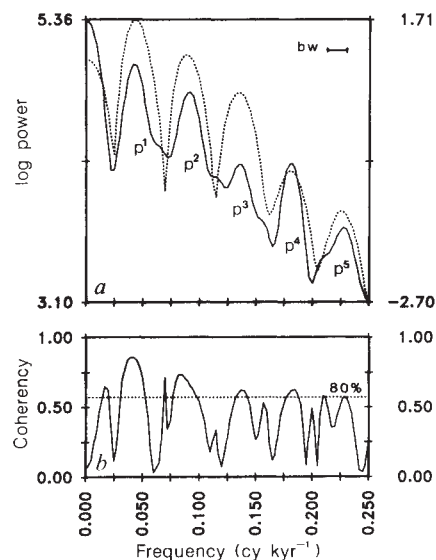
their counterparts generated from the orbital precession index $e \cdot \sin \tilde{\omega}$ (ref. 9). That is, their relationships are statistically significant at the 80% level. The record is too short to test reliably for the presence of the ~100 kyr eccentricity period. Additional cross-spectral tests reveal no evidence for coherent spectral power at the Earth's obliquity period (41 kyr), or its harmonics. Thus, essentially all the down-core variations of *Melosira* in core V30-40 are related in some way to precession of the Earth's equinoxes.

It has been suggested that low summer solar radiation in the Northern Hemisphere (that is, during periods of high precession index $e \cdot \sin \tilde{\omega}$, defined as Earth-Sun distance at the summer solstice⁹) would suppress the monsoonal circulation and result in arid episodes in Eurasia and tropical Africa^{7,10-13}. The clear presence of a precessional signal in our *Melosira* data supports theories calling for precessional control of climates in tropical Africa.

The record of *Melosira* accumulation rates in V30-40, however, leads the orbital precession index ($e \cdot \sin \tilde{\omega}$), or low Northern Hemisphere summer solar radiation in the 23 kyr precessional band, by about 35° (2-3 kyr). This lead is not an artefact of the SPECMAP⁵ timescale. For comparison, the same *Melosira* data expressed on the earlier Morley and Hays¹⁴ timescale lead precession by about 80° (5 kyr). *Melosira* maxima are approximately in phase with Northern Hemisphere solar radiation minima in April to May (SPECMAP⁵ timescale) or March to April (Morley and Hays¹⁴ timescale). This could suggest sensitivity of monsoonal climates to solar radiation variations in spring, rather than summer.

Alternatively, the phase difference between Northern Hemisphere summer insolation and *Melosira* may be an artefact of the recording of lake levels by *Melosira*. A linear response of *Melosira* accumulation rates to low-latitude solar radiation would yield a spectrum dominated by the ~23 kyr precessional peak. The presence of coherent higher-frequency harmonics of precession argues that the response of the *Melosira* record to climate change is highly nonlinear. That is, the record is 'spiky', although variations in the forcing are smooth. A possible reason for nonlinear recording of climate by *Melosira* is depletion of the supply of diatomaceous sediments, either by complete erosion or by the development of deflation-resistant crusts or soils on exposed lake beds following desiccation¹⁵. The result would be a relatively brief input of *Melosira* to ocean sediments during or immediately following desiccation of lakes, and prior to maximum aridity (Fig. 3). The apparent lead of *Melosira*

Fig. 2 *a*, The log power spectra of *Melosira* accumulation rates (solid, left axis) and orbital precession ($e \cdot \sin \tilde{\omega}$, ref. 9) plus its second through fifth harmonics (dashed, right axis) demonstrate shared variance in the dominant frequency bands ($p^1 = 23$, $p^2 = 11.5$, $p^3 = 7.4$, $p^4 = 5.5$ and $p^5 = 4.4$ kyr). The records are too short to be certain of the match of frequencies at longer periods (near 100 kyr). Bandwidth is indicated by bw. *b*, Coherency spectrum of *Melosira* accumulation rates versus orbital precession plus harmonics confirms the statistical significance of the relationship between the time series in all of the precession-related frequency bands (dashed line indicates 80% confidence level). Thus, essentially all of the variations in the *Melosira* record are related to orbital precession.



maxima ahead of Northern Hemisphere summer solar radiation minima may thus be an artefact of erosional and redepositional processes. Similarly, the short duration of *Melosira* maxima does not necessarily indicate that arid intervals in tropical Africa were brief. Our results are consistent with aridity maxima in equatorial Africa during (or even somewhat later than¹⁶) minima in Northern Hemisphere summer solar radiation.

This conceptual model does not account for some of the smaller features of the *Melosira* time series. Secondary *Melosira* maxima in peak glacial stages 2, 4, and 6 of V30-40 appear to be synchronous with major *Melosira* maxima in core V30-49, at a site under the summer dust plume at 18°26' N, 21°05' W (ref. 1). The secondary maxima at these times in V30-40 may indicate greater influence of summer-plume dust near the equator during times of increased global ice volume, due either to southward expansion of the Sahara or to southward displacement of summer wind systems at the glacial maximum. This concept of equatorward compression of climatic belts during glaciation has been supported by Nicholson¹⁷, but contested by Tetzlaff *et al.*¹⁸. A test of this hypothesis will come from collection

of additional data on the latitudinal extent of past eolian contributions to marine sediments.

A complete understanding of *Melosira* in deep-sea sediments awaits additional data collection and the development of more realistic models of *Melosira* transport to and preservation in oceanic sediments. At present, however, the results strongly support hypotheses calling for direct modulation of tropical (monsoonal) climates in tropical Africa by low-latitude solar radiation variations, which are dominated by the ~23 kyr precessional period. The apparent lack of a 41 kyr obliquity period in the *Melosira* data suggests that changes in high-latitude solar radiation and feedback from global ice volume changes do not contribute strongly to climate change in this area. This confirms general circulation model predictions¹⁹ that suggest changes in tropical (monsoonal) precipitation are much more sensitive to changes in seasonal radiation forcing than to changes of glacial-age boundary conditions.

We thank N. G. Pisias, W. F. Ruddiman and T. C. Janecsek for reviews of the manuscript and J. E. Kutzbach, W. L. Prell, and F. A. Street-Perrott for discussions. E.M.P. acknowledges financial support from the South African CSIR/Foundation for Research Development and NSF grant OCE86-08328. A.C.M. acknowledges support from NSF grant ATM83-19371 (project SPECMAP). L-DGO contribution no. 4086.

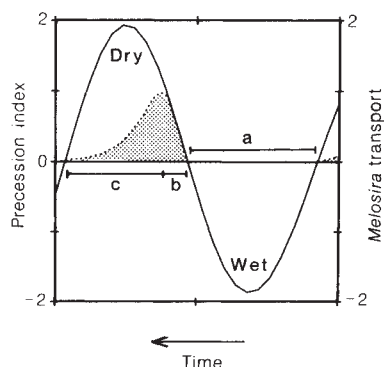


Fig. 3 A cartoon demonstrates how maximum *Melosira* input to the ocean (shaded) may predate maximum aridity over an approximately 23 kyr precessional cycle. Lake levels are high during time interval *a*, and diatomaceous sediments cannot be eroded. Lake levels fall during time interval *b*, and a large area of sediment is exposed and eroded. With continued fall in lake levels (time interval *c*), previously exposed deposits have been depleted. Smaller areas of sediment are exposed, and formation of soils or crusts on dry lake beds inhibit continued eolian transport despite continuing aridity.

Received 22 September 1986; accepted 6 January 1987.

1. Pokras, E. M. & Mix, A. C. *Quat. Res.* **24**, 137-149 (1985).
2. Stabell, B. *Marine Geol.* **75**, 305-323 (1986).
3. Kalu, A. E. in *Saharan Dust: Mobilization, Transport, Deposition* (ed. Morales, C.) 95-118 (Wiley, Chichester, 1979).
4. Prospero, J. M. *Spec. Pap. geol. Soc. Am.* **186**, 71-86 (1981).
5. Imbrie, J. *et al.* in *Milankovitch and Climate, Part 1* (eds Berger, A. L. *et al.*) 269-305 (Riedel, Dordrecht, 1984).
6. Curry, W. B. & Lohmann, G. P. in *The Carbon Cycle and Atmospheric CO₂: Natural Variations Archean to Present* (eds Sunquist, E. T. & Broecker, W. S.) 285-301 (AGU, Washington DC, 1985).
7. Kutzbach, J. E. & Street-Perrott, F. A. *Nature* **317**, 130-134 (1985).
8. Jenkins, G. M. & Watts, D. G. *Spectral Analysis and its Applications* (Holden-Day, San Francisco, 1968).
9. Berger, A. L. *J. atmos. Sci.* **35**, 2362-2367 (1978).
10. Bernard, E. A. *Acad. R. Sci. Outre-Mer Classe Sci. Nat. Med. Nouvelle Serie XII*, Fasc 1, 1-232 (1962).
11. Kutzbach, J. E. in *Variations in the Global Water Budget* (eds Street-Perrott, F. A., Beran, M. & Ratcliffe, R. A. S.) 371-389 (Reidel, Dordrecht, 1983).
12. Rossignol-Strick, M. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **49**, 237-263 (1985).
13. Prell, W. L. & Van Campo, E. *Nature* **323**, 526-528 (1986).
14. Morley, J. J. & Hays, J. D. *Earth planet. Sci. Lett.* **53**, 279-295 (1981).
15. Talbot, M. R. in *The Sahara and the Nile* (eds Williams, M. A. J. & Faure, H.) 37-62 (Balkema, Rotterdam, 1980).
16. Short, D. A. & Mengel, J. G. *Nature* **323**, 48-50 (1986).
17. Nicholson, S. E. *Nature* **296**, 779 (1982).
18. Tetzlaff, G., Wolter, K. & Sarythein, M. *Nature* **296**, 779-780 (1982).
19. Kutzbach, J. E. & Guetter, P. J. *J. atmos. Sci.* **43**, 1726-1759 (1986).

Latitudinal shift of Pacific hotspots during the late Cretaceous and early Tertiary

William W. Sager* & Ulrich Bleil†

* Department of Oceanography, Texas A&M University, College Station, Texas 77843, USA

† Fachbereich Geowissenschaften Universität Bremen, 2800 Bremen 33, FRG

Two frames of reference, the geomagnetic field and hotspots, have been widely used to measure the past motions and interactions of lithospheric plates. Several studies have found that these reference frames have undergone as much as 19–22° of relative motion, often called true polar wander (TPW), in the past 200 Myr^{1–4}. Moreover, it also appears that this offset did not accumulate at a uniform rate. Some evidence suggests that spurts of TPW can occur at rates as high as 1–2° per Myr^{3–5}. Rates even an order of magnitude less could pose serious problems for tectonic studies based on the hotspot reference frame because velocities of this magnitude cannot be considered insignificant compared to those of the plates. Unfortunately, the imprecision of most palaeomagnetic data is such that large error bounds are common for many estimates of TPW rates. In this report, we present a record of TPW derived from Pacific palaeomagnetic data that are relatively well constrained in age and palaeolatitude. They show that offset between the two reference frames accumulated at ~0.3° per Myr⁻¹ from late Cretaceous to early Tertiary.

Palaeomagnetic samples locate a given site with respect to the ancient geomagnetic field which is assumed to be a geocentric dipole aligned along the spin axis. Dated volcanic chains record the movement of the plates over mantle plumes, considered by many to constitute an absolute reference frame^{1,6}. Most authors

agree that 5–12° of relative motion has occurred between these reference frames since the late Cretaceous, with the spin axis moving away from the Pacific Ocean^{1,3–5,7}. However, previous analyses were unable to resolve the timing and rate of the offset⁷. Investigations based on globally distributed palaeomagnetic data sets, back-tracked in the hotspot reference frame, have given seemingly disparate results. Livermore *et al.*⁸ concluded that an insignificant amount of offset of the two reference frames has occurred in the past 35 Myr. However, others^{1–3} found that between 7–10° of TPW has accumulated in the past 45–50 Myr. The source of this discrepancy is unclear; however, the studies by Livermore *et al.* included data from the Pacific plate, whereas the others did not. Perhaps these data tend to 'cancel' some of the apparent TPW observed with less complete data sets.

Palaeomagnetic data from the Pacific plate are particularly useful for the study of such phenomena because it has experienced rapid northward translation since the Cretaceous. Furthermore, sediment facies data from this plate are available for an independent measurement of changes in palaeolatitude. A recent review of palaeomagnetic data from the Hawaiian-Emperor chain concluded that little relative motion has occurred between the Pacific hotspots and the spin axis since the Eocene⁹. However, the 27.6° N palaeolatitude of Suiko Guyot¹⁰, in the Emperor chain, was used to infer that relatively rapid TPW had occurred in the late Cretaceous and early Tertiary⁷. Unfortunately, available Pacific palaeomagnetic data clustered around Eocene and Maastrichtian times, leaving a critical gap in the Palaeocene which made delineation of the rate of TPW difficult. Data from this gap have recently been obtained from a study of Deep Sea Drilling Project (DSDP) cores from holes 577 and 577A, located in the western Pacific¹¹.

To examine the offset of the palaeomagnetic and hotspot reference frames in the late Cretaceous and early Tertiary, we have assembled palaeomagnetic data from the Pacific (Table 1) which are well constrained in both age and palaeolatitude. We back-tracked these in time, using Euler poles of rotation describ-

Table 1 Late Cretaceous–early Tertiary palaeomagnetic data

ID	Name	Latitude (°N)	Longitude (°E)	Age (Myr)	Age type	Data type	Palaeolatitude (observed)	95% Confidence	Palaeolatitude (calculated)	δL	Palaeomagnetism ref.	Age ref.
A	S6824	-2.0	181.2	32–36	F	SC	-15.6	6.4	-12.2	-3.4	*	*
B	S6824	-2.0	181.2	38–39	F	SC	-10.6	6.8	-13.6	3.0	12	12
C	Abbott	31.8	174.3	38.5 ± 2.5	A	SM	17.5	7.0	20.3	-2.8	9	31
D	Colahan	31.0	176.0	38.5 ± 1.5	A	SM	18.8	7.0	19.5	-0.7	18	31
E	Stanley	8.2	198.1	39.3 ± 1.1	A	SM	-5.3	7.0	-3.0	-2.3	27	27
F	Willoughby	7.8	198.1	same as E	—	SM	-3.9	7.0	-3.4	-0.5	27	27
G	M7039	-2.5	186.7	39–41	F	SC	-16.6	5.2	-14.4	-2.2	12	12
H	K781012	11.4	191.8	48–50	F	SC	-3.1	4.8	-4.4	1.3	12	12
I	DSDP 577, 577A	32.4	157.7	50–53	M	SC	16.8	5.0	15.0	1.8	11	11
J	DSDP 577, 577A	32.4	157.7	54–56	M	SC	15.5	5.2	12.9	2.6	11	11
K	DSDP 577, 577A	32.4	157.7	56–61	M	SC	14.5	5.0	10.8	3.7	11	11
L	DSDP 577, 577A	32.4	157.7	61–65	M	SC	14.9	4.2	8.3	6.6	11	11
M	DSDP 433C	44.8	170.0	64.7 ± 1.1	A	BC	26.1	8.6	19.2	6.9	11	32
N	DSDP 199	13.5	156.2	62–67	M	SC	-6.0	8.2	-11.6	5.6	†, 17	33
O	Paumakua	24.9	202.9	65 ± 5	A	SM	3.9	5.7	3.2	0.7	18	34, 18
P	GPC3	30.0	202.2	>65	F	SC	11.1	7.0	8.3	2.8	28	28
Q	DSDP 577, 577A	32.4	157.7	65–67	M	SC	12.0	4.4	6.2	5.8	11	11
R	DSDP 577, 577A	32.4	157.7	67–68	M	SC	13.6	3.6	5.0	8.6	11	11
S	DSDP 315A	4.2	201.5	66–74	F	SC	-9.7	4.2	-20.0	10.3	29	29
T	DSDP 585, 585A	13.5	156.8	66–74	F	SC	-1.3	4.4	-15.6	14.3	30	30
U	Wagman	-7.5	208.5	71.9 ± 1.4	K	SM	-22.8	5.7	-29.7	6.9	27	27
V	Bach Ridge	26.6	200.1	73.6 ± 1.4	K	SM	9.7	5.2	-1.3	11.0	18	34, 18
W	DSDP 165	8.2	195.1	70–78	M	SC	-9.7	5.2	-20.5	10.8	†, 17	35
X	H11	26.5	182.2	74.3 ± 4	K	SM	4.6	5.7	-5.4	10.0	18	36
Y	H13	28.1	188.8	74.5 ± 3.5	K	SM	7.0	5.7	-2.9	9.9	18	36
Z	Haydn	26.6	198.7	75 ± 3	A	SM	7.0	5.7	-2.5	9.5	18	34, 18

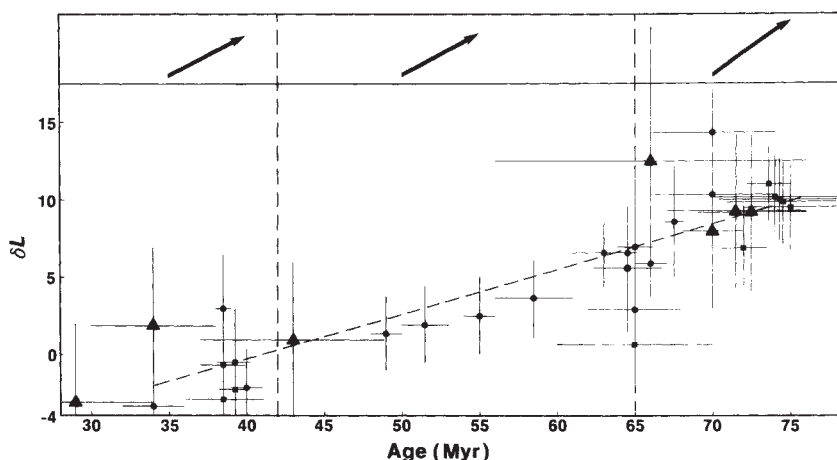
Data type: SC, sediment core; SM, seamount palaeomagnetism; BC, basalt core.

Age type: F, fossil; A, Ar⁴⁰/Ar³⁹; M, magnetic reversals; K, K-Ar.

* Unpublished data (F. Theyer, personal communication).

† Unpublished data (B. H. Keating, personal communication).

Fig. 1 Palaeolatitude data plotted in the hot-spot reference frame. The mean age of each sample in Myr is plotted along the horizontal axis, whereas the difference between the observed palaeolatitude and palaeolatitude predicted from a model of the motion of the Pacific plate versus the hotspots, δL , is plotted on the vertical axis. The error bars correspond to the standard deviation in palaeolatitude and age. For those data with no standard deviation given for the age, the error bars are the given age range. Triangles, equatorial transit data; circles, DSDP and piston core data; squares, seamount palaeomagnetic data. Dashed sloping line, the regression line fit to the data. Vertical lines, the ages at which the rotation poles change in the plate motion model. Top arrows, 'movement' on the figure caused by an error of 5 Myr in determining the age of any datum.



ing the motion of the Pacific plate with respect to the hotspots. A latitude error, δL , was calculated by subtracting the palaeolatitude predicted by the hotspot rotation poles from the observed palaeolatitude. This quantity is plotted against age in Fig. 1. A decrease of δL corresponds to a southward movement of the Pacific hotspots. Our study was limited to the interval 30–75 Myr, because several other studies have examined the implications of post-Eocene Pacific palaeomagnetic data^{9,12,13} and because the motion of the Pacific with respect to the hotspots is not well constrained for times before 75 Myr.

We chose to use the hotspot stage poles of Duncan and Hargraves¹⁴ for back-tracking the data. Their single pole describing the 0–42 Myr interval (corresponding to the Hawaiian chain) is similar to those determined by others for the same time span. However, they used two poles to describe hotspot motion in the Palaeocene and Maastrichtian (corresponding to the Emperor chain), rather than the single pole used by most authors. Their pole for 42–65 Myr relies mostly on the dated southern Emperor seamounts for constraint of the rotation rate, whereas the 65–74 Myr pole uses mainly ages from the central and southern Line Islands seamounts for constraint. Use of the most recent age data available make this model of Pacific-hotspot relative motion the best for our purposes; nevertheless, our conclusions do not appear to be dependent on the plate motion model used.

The palaeomagnetic data used in this study are primarily of two types, DSDP and piston core measurements and the inversion of seamount magnetic anomalies. Adequate discussions of these methods of obtaining palaeomagnetic data are not possible here, for reviews see refs 15–17. A few comments about the data are appropriate, however. We attempted to select only the most reliable palaeomagnetic data for this study. All the cores have been partially demagnetized to remove secondary magnetizations. Furthermore, using the method of Cox and Gordon¹⁵, these data have been corrected for systematic errors inherent in inclination-only data. The seamount palaeomagnetic data all have high goodness-of-fit ratios, a parameter measuring the match of the observed and modelled magnetic anomalies.

For seamount palaeomagnetic data, the assumption is made that induced and viscous contributions to the measured magnetization directions are negligible. Given that the removal of such spurious components from a seamount's magnetization *in situ* is impossible, this may seem an alarming assumption. However, Koenigsberger ratios of most oceanic basalts are large. Furthermore, comparison of seamount and nonseamount palaeomagnetic data in this and other studies^{17,18} shows that they agree remarkably well; thus, bias due to secondary magnetizations appears small.

The palaeomagnetic data in Fig. 1 show an obvious, nearly linear trend. At ~40 Myr, the observed palaeolatitudes agree with the hotspot palaeolatitudes. However, at 75 Myr an offset between the two reference frames of 9–12° is seen. In the inter-

vening time, the offset appears to decrease in a more or less linear manner, suggesting TPW with the Pacific hotspots moving southward.

Because this study examines only data from the Pacific, care must be exercised not to misinterpret other phenomena as TPW. Two possible culprits are long-term non-dipole components and errors in the Pacific-hotspot rotation poles. A number of studies have shown that, despite time-averaging, significant non-dipole components can remain in the geomagnetic field^{19,20}. These can rarely be recognized solely with the use of palaeomagnetic data from a limited geographical area; however, an additional constraint on the location of the spin axis comes from equatorial transits observed in Pacific DSDP cores. The crossing of the Equator by a DSDP site is often recorded by a characteristic suite of sediments generated by the equatorial zone of high biological productivity²¹. This yields a datum akin to a palaeomagnetic latitude, but unaffected by non-dipole fields. Pacific equatorial transit data of late Cretaceous and early Tertiary age, listed elsewhere⁹, have been back-tracked and plotted in Fig. 1. Their excellent agreement with the palaeomagnetic data leads us to believe that TPW is the primary cause of the observed offset. As other studies have found evidence for non-dipole fields during this period^{4,22}, we cannot rule out their contribution. The two phenomena probably occur simultaneously and are difficult to separate.

The offset between the reference frames does not appear to be the result of errors in the rotation poles used to describe Pacific-hotspot motion. Uncertainties of several degrees might be expected in the determination of the location of a hotspot due to perturbations of its extrusions by lithospheric motion and structure²³. Furthermore, the depths and ages of the seamounts used to derive the Euler poles used in this study are well enough constrained that systematic errors much greater than 2–3° are not expected. Therefore, a large systematic offset, such as that observed in Fig. 1, seems unlikely to result solely from errors in the rotation poles.

As noted above, the trend of the offset between the hotspots and the palaeomagnetic axis is approximately linear. A line fit to all the data in Table 1 yields a slope of 0.27° Myr⁻¹ and a correlation coefficient of 0.89. Thus the data record an event of TPW that began some time before 75 Myr and continued at a relatively uniform rate until the early Tertiary. Note that this rate is a lower bound on the true rate because only changes in latitude were examined. However, as other studies indicate that the spin axis was moving almost directly away from the Pacific^{1,3–5,7}, this value is probably close to the true rate.

The date at which the linear trend crosses zero, 41 Myr, is suggestive. This is very nearly the age, 42.5 Myr, determined for the bend in the Hawaiian-Emperor chain²⁴. Although this bend is usually attributed to a change in plate motion, this correspondence suggests that it may also be in part due to a change in the motion of the hotspot itself. Moreover, our findings imply

that nearly half of the latitudinal range of the Emperor Seamounts may be a result of TPW.

Hotspots and subducting lithospheric slabs have been suggested as the two major mantle mass anomalies that control TPW²⁵. Moreover, the pull of subducting slabs is generally recognized as the primary driving force acting on plates with trench boundaries. Thus the Hawaiian-Emperor bend, probably caused by a change in the configuration of the subducting slabs on its boundaries²⁶, might be expected to be intimately related to changes in TPW.

Received 30 October 1986; accepted 19 January 1987.

1. Morgan, W. J. *The Sea* Vol. 7, 443-487 (Wiley, New York, 1981).
2. Harrison, C. G. A. & Lindh, T. *Nature* **300**, 251-252 (1982).
3. Andrews, J. A. *J. geophys. Res.* **90**, 7737-7750 (1985).
4. Livermore, R. A., Vine, F. J. & Smith, A. G. *Geophys. J. R. astr. Soc.* **79**, 939-961 (1984).
5. Gordon, R. G. *Geophys. Res. Lett.* **10**, 709-712 (1983).
6. Duncan, R. A. *Tectonophysics* **74**, 29-42 (1981).
7. Gordon, R. G. & Cape, C. D. *Earth planet. Sci. Lett.* **55**, 37-47 (1981).
8. Livermore, R. A., Vine, F. J. & Smith, A. G. *Geophys. J. R. astr. Soc.* **73**, 153-171 (1983).
9. Sager, W. W. *J. geophys. Res.* **89**, 6271-6284 (1984).
10. Kono, M. *Init. Rep. DSDP* **55**, 737-752 (1980).

11. Bleil, U. *Init. Rep. DSDP* **76**, 441-458 (1985).
12. Epp, D., Sager, W. W., Theyer, F. & Hammond, S. R. *Nature* **303**, 318-320 (1983).
13. Hammond, S. R., Epp, D. & Theyer, F. *Nature* **278**, 309-312 (1979).
14. Duncan, R. A. & Hargraves, R. B. *Mem. geol. Soc. Am.* **162**, 81-93 (1984).
15. Cox, A. & Gordon, R. G. *Rev. Geophys. Space Phys.* **22**, 47-72 (1984).
16. Harrison, C. G. A., Jarrard, R. D., Vacquier, V. & Larson, R. L. *Geophys. J. R. astr. Soc.* **42**, 859-882 (1975).
17. Sager, W. W. thesis, Univ. Hawaii (1983).
18. Sager, W. W. *Tectonophysics* (in the press).
19. Coupland, D. H. & Van der Voo, R. *J. geophys. Res.* **85**, 3529-3548 (1980).
20. Merrill, R. T. & McElhinny, M. W. *Rev. Geophys. Space Phys.* **15**, 309-323 (1977).
21. van Andel, T. H., Heath, G. R. & Moore, T. C. *Mem. geol. Soc. Am.* **143** (1975).
22. Idnurm, M. *Geophys. J. R. astr. Soc.* **83**, 419-433 (1985).
23. Epp, D. *J. geophys. Res.* **89**, 11273-11286 (1984).
24. Clague, D. A. & Dalrymple, G. B. *Geophys. Res. Lett.* **2**, 305-308 (1975).
25. Jurdy, D. M. *J. geophys. Res.* **88**, 6395-6402 (1983).
26. Gordon, R. G., Cox, A. & Harter, C. E. *Nature* **274**, 752-755 (1978).
27. Sager, W. W. & Keating, B. H. *J. geophys. Res.* **89**, 11135-11151 (1984).
28. Prince, R. A., Heath, G. R. & Komins, M. *Bull. geol. Soc. Am.* **91**, 1789-1835 (1980).
29. Cockerham, R. S. & Jarrard, R. D. *Init. Rep. DSDP* **33**, 631-647 (1976).
30. Ogg, J. R. *Init. Rep. DSDP* **89**, 629-646 (1986).
31. Duncan, R. A. & Clague, D. A. *Eos* **65**, 1076 (1984).
32. Dalrymple, G. B., Lanphere, M. A. & Clague, D. A. *Init. Rep. DSDP* **55**, 659-676 (1980).
33. Scientific Staff DSDP, *Init. Rep. DSDP* **20**, 65-86 (1973).
34. Sager, W. W. & Pringle, M. S. *Seamounts, Islands, and Atolls*, *Am. Geophys. Un. Monogr. Ser.* (in the press).
35. Scientific Staff DSDP, *Init. Rep. DSDP* **17**, 47-102 (1973).
36. Garcia, M. O., Grooms, D. G. & Naughton, J. J. *Lithos* (in the press).

Authigenic magnetite formation in suboxic marine sediments

Robert Karlin*, Mitchell Lyle† & G. Ross Heath*

* School of Oceanography, University of Washington, Seattle, Washington 98195, USA

† College of Oceanography, Oregon State University, Corvallis, Oregon 97331, USA

The resolution and reliability of magnetostratigraphy and reconstructed time series of geomagnetic field behaviour depend critically on where remanence is acquired and fixed in the sediment column and whether the magnetization is altered chemically after deposition. If authigenic magnetic minerals are formed at depth or if the magnetic carriers are changed after deposition, then the nature and timing of magnetic events can be affected by depth offsets in remanence acquisition and mixing of detrital and authigenic signals. Using palaeo- and rock-magnetic and sediment geochemical analyses, we have studied how early diagenesis affects the magnetic properties of suboxic hemipelagic sediments. Here, we report evidence of the formation of fine-grained authigenic magnetites near the commonly observed brown-tan-green colour boundary, which marks the transition from Fe-oxidizing to Fe-reducing conditions. We propose that biogenic magnetite, produced by magnetotactic bacteria, forms as part of the microbially mediated sequence of reactions involved in the oxidation of organic matter. The magnetite is created as a metabolic by-product of the microorganisms' use of iron redox transitions as a source of energy. Active magnetite formation appears to be restricted to a zone between the levels of nitrate and iron reduction.

Early diagenesis of iron in sediments occurs as a part of the progressive oxidation of organic matter through a sequence of microbially mediated metabolic reactions proceeding from O₂, NO₃, Mn, Fe and SO₄ reduction to fermentation and CH₄ production¹⁻⁴. Suboxic marine sediments commonly show a near surface brown-tan-green colour zonation related to these diagenetic reactions. Oxidation (by O₂), nitrate reduction and MnO₂ precipitation are usually confined to the surficial brown oxidized zone, whereas Mn⁴⁺ reduction, and possibly Fe²⁺ oxidation, occur in the transitional tan layer. Iron reduction begins at the top of the underlying green zone⁵⁻⁷. The extent of reactions and the depth of occurrence in a given sedimentary column depend primarily on the availability of oxidants, the input of metabolizable organic matter and the sediment accumulation rates⁸⁻⁹.

Magnetotactic bacteria have been reported from bogs, lakes and open marine environments, but the environmental factors controlling their distributions are not well known¹⁰⁻¹³. They appear to be microaerophilic iron reducers, rather than obligate anaerobes, and may require a hydrous ferric oxide precursor (ferrihydrite) to form relatively pure single domain (SD) magnetite¹⁴⁻¹⁵. Magnetites with similar properties have been reported from ancient sediments^{16,17} and recently in pelagic South Atlantic sediments¹⁸, leading to speculation that biogenic magnetites may be an important part of the palaeomagnetic record.

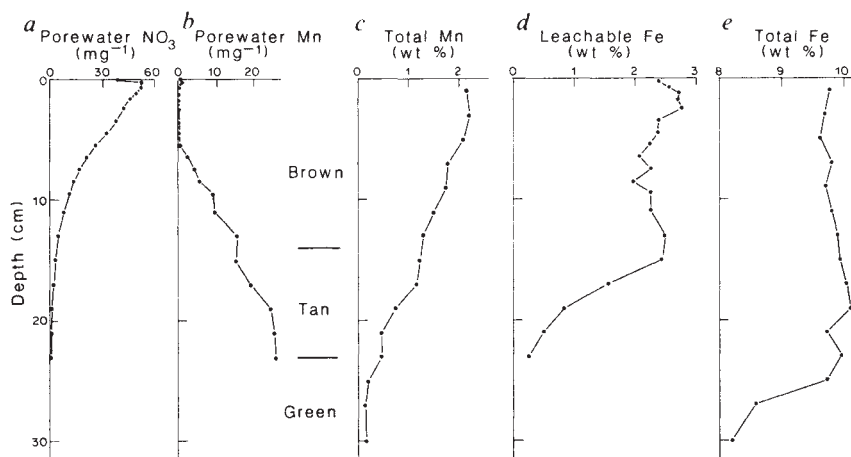
Using integrated rock-magnetic, chemical and mineralogical techniques, we have studied several cores from MANOP Site M in the eastern equatorial Pacific (8.47°S N, 104°4' W) to assess the effects of iron diagenesis on the formation and destruction of magnetic minerals. Our objectives were to determine the conditions in the marine environment in which magnetic minerals are formed or altered and the factors that affect their preservation in the sediment column.

The sediments of Site M, located 25 km from the East Pacific Rise, receive biogenic, continental and minor, but significant, hydrothermal inputs^{19,20}. The sediments have a distinctive stratigraphy with a brown-tan-green sequence less than 5 m thick overlying ~2 m of basal yellowish brown-orange metalliferous clay. Sedimentation rates are 1.0-1.25 cm kyr⁻¹, based on ²³⁰Th and ¹⁸O stratigraphy^{19,21}.

In box core M25 from the Pluto III expedition, porewater profiles (S. Emerson, personal communication) show nitrate rapidly decreasing from just below the surface to the brown-tan transition at 14 cm (Fig. 1a). Porewater Mn is zero from 0-5 cm then increases linearly until the tan/green colour boundary at 23 cm (Fig. 1b). These data and the solid phase Mn profile (Fig. 1c) suggest manganese reduction and remobilization in the tan zone and oxidation and precipitation, presumably as MnO₂, in the brown surface layer at about 5 cm depth. Although porewater iron was not measured in this core, analyses on a nearby core show negligible Fe(II) in the brown and tan zones, and an increase with depth starting near the top of the green layer, presumably marking the onset of Fe reduction²². Based on ²³⁰Th and ²²⁶Ra profiles, bioturbation is limited to the upper 5 cm of the sediments²¹.

The bulk chemistry of the M25 sediments was determined by both X-ray fluorescence and atomic absorption to define any downcore changes in provenance. Biogenic carbonate increases from 14% at 0-8 cm to 30-34% below 18 cm. On a carbonate-free basis, the terrigenous elements Al, Mg, Ti and K are constant from 0-25 cm, then increase by about 20% below 25 cm, suggesting a slightly higher contribution of detrital aluminosilicates at depth. In contrast, total iron (Fig. 1e) is relatively constant to

Fig. 1 Porewater profiles of; *a*, nitrate, *b*, manganese and solid phase profiles of; *c*, total manganese, *d*, iron extractable in 1 M hydroxylamine hydrochloride-sodium citrate (*pH* 5.0) and *e*, total iron for box core M25 (8.47°S'N, 104°4'W) from MANOP Site M. Solid phase elements are given on a salt and carbonate-free basis.



25 cm, then decreases by about 15% downcore. This decrease could be due either to a lesser hydrothermal input ~21–25 kyr ago, or to remobilization at depth and precipitation at the top of the Fe-reduction zone at 23 cm.

To define the iron-bearing phases better, the sediments were subjected to a sequential extraction procedure with buffered acetic acid (*pH* 5), sodium dodecyl sulphate (SDS), and buffered (*pH* 5.0) hydroxylamine HCl (HH), and analysed by atomic absorption spectroscopy^{20,23}. The acetic acid treatment removes carbonates and sorbed constituents. The SDS leach extracts metals that are organically chelated. The HH treatment removes amorphous and poorly crystalline Fe and Mn oxides and oxyhydroxides, but does not attack magnetite, based on experiments with pure minerals.

In M25, the HH treatment was the only leach that dissolved significant amounts of iron. In Fig. 1*d*, HH-extractable iron is high (2.4–3.0%) and variable in the brown oxidized layer from 0–15 cm, then declines rapidly from ~2.4% to 0.3% in the Mn reduction zone. These results suggest that a leachable iron fraction, which is relatively constant in the nitrate reduction zone, is being transformed to a more stable form in the transitional tan zone from 14–23 cm.

The natural remanent magnetization (NRM) intensity in sediments depends on the type of magnetic carriers, their concentration, size and shape, the particle alignment efficiency and the geomagnetic field. In carbonate-rich sediments, the NRM often mirrors the carbonate content, and, to a lesser extent, the water content²⁴. This dilution effect can be mitigated by normalizing the NRM to the non-carbonate fraction on a dry-weight basis.

When corrected for salt and carbonate contents, NRM intensities in M25 (Fig. 2*a*) are essentially constant in the brown oxidized layer from 0–14 cm. In the Mn reduction zone, intensities increase, reaching a maximum just below the tan-green colour change at 23 cm. NRM stabilities (Fig. 2*b*), as measured by the median demagnetizing field (MDF), show the same trends with relatively high values in the brown and tan zones and a peak at the top of the iron reduction zone. The high MDFs are mostly due to a magnetically hard component that cannot be demagnetized at 1 kOe alternating field (AF). In Fig. 3, a downcore profile of NRM-normalized magnetizations (J/J_0) at different AF levels shows the presence of two types of magnetic carriers with different coercivities. A softer component with coercivities <200 Oe AF is dominant in the brown zone, and remains constant or slightly decreases in the tan and green layers. A hard component with coercivities >300 Oe AF increases downcore to a peak at the tan-green boundary then declines in abundance. In magnetite, such high coercivities are usually diagnostic of fine-grained SD or pseudo-single domain (PSD) particles²⁵.

The NRM intensity and MDF peaks suggest that the layer overlying the iron reduction zone contains an enhanced contri-

bution of fine-grained magnetic material. The NRM increase could have been due to a change in source input or to the formation of an authigenic magnetic phase at depth. NRM directions provide indirect evidence that the subsurface NRM peak is of authigenic origin. NRM inclinations (Fig. 4*a*), demagnetized at 200 Oe, show considerable variability above 20 cm, both within and among horizons, consistent with the slow sedimentation rate and the presence of active bioturbation. However, below 20 cm, the directional scatter decreases and core inclinations fall closer to the present day inclination (22°) than to the expected inclination due to a geocentric axial dipole (17°). The azimuthally unoriented declinations (Fig. 4*b*) in the tan and upper green zones also cluster about a different direction than in the brown layer.

A comparable pattern of variability was seen in each of the samples upon AF demagnetization. Using β_{95} (analogous to the Fisherian α_{95}) as a measure of directional scatter for 6–7 steps from 0–400 Oe AF for each sample, Fig. 4*c* shows that this parameter drops from >4° in the brown zone to <2° in the tan and green layers, then increases again at the base of the core. This behaviour suggests that the magnetic carriers in the brown zone contain multiple components of remanence, presumably acquired at different times, whereas the remanence in the zone of high NRM intensity is stable about a single direction that is close to the present day field. The relatively large scatter at the top of the core may be partially due to bioturbation.

To identify the magnetic carriers, magnetic separates from six levels of M25 were analysed by X-ray diffraction using a Scintag

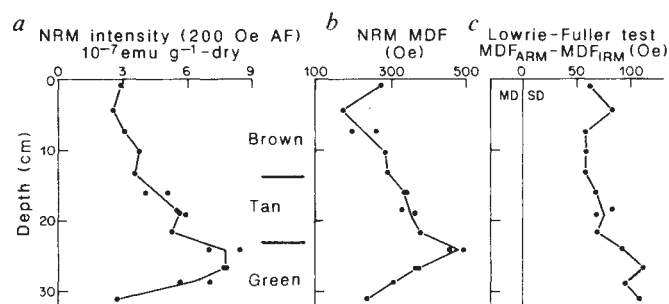


Fig. 2 Magnetic properties of core M25. *a*, NRM intensities, demagnetized at 200 Oe AF and normalized to unit weight (salt and carbonate free), and *b*, NRM MDFs, showing maxima at the tan-green colour boundary, coincident with a loss of leachable iron oxides. *c*, Profile of modified Lowrie-Fuller test comparing differences in MDFs of ARM ($h(d.c.)=0.5$ Oe, $H(AF)=1$ kOe) and SIRM (10 kOe). To compare the same coercivity spectrum, the remanence remaining after 1 kOe AF was subtracted from each treatment. The magnetites in this core show SD-PSD behaviour throughout the core.

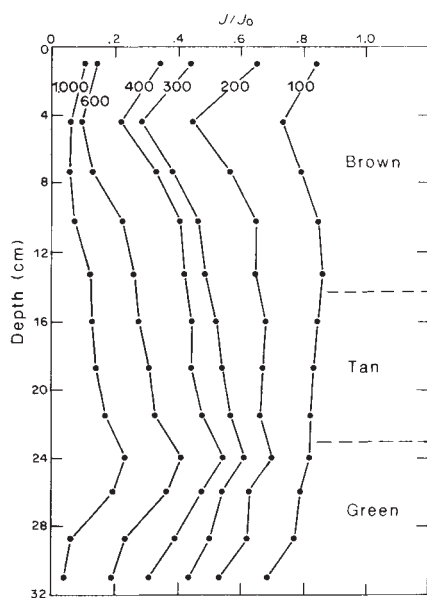


Fig. 3 Downcore profile of magnetizations normalized to initial NRM (J/J_0) for different AF demagnetization levels. A magnetically soft component is dominant in the oxidized brown zone, and remains relatively constant downcore. A high coercivity component (presumably SD or PSD magnetite) increases in abundance to the tan-green boundary marking the onset of iron reduction, then decreases as the fine-grained magnetic carriers begin to be dissolved.

powder diffraction unit (Cu radiation with a Ni monochromator). In all samples, magnetite was the only ferrimagnetic mineral identified. Based on at least four reflections per diffractogram, the unit cell parameter for the magnetite averaged $8.40 \pm 0.01 \text{ \AA}$, indicating almost pure Fe_3O_4 . Isothermal remanence (SIRM) and saturation magnetization (J_s) acquisition curves reach saturation by 3 kOe which is also suggestive of magnetite.

The domain states and concentrations of the magnetites in M25 were determined from hysteresis loop parameters obtained on a vibrating sample magnetometer. Ratios of saturation remanence/saturation magnetization and remanent coercivity/coercivity for 10 samples fall within a narrow range ($J_{rs}/J_s = 0.28 \pm 0.01$; $H_{rc}/H_c = 2.2 \pm 0.1$), suggesting that the overall magnetite distribution is dominated by PSD-sized grains or a mix of SD and PSD. These results are consistent with a modified Lowrie-Fuller test based on the difference in MDF between anhysteretic (ARM) remanence and SIRM^{26,27}. In Fig. 2c, the positive values of $L_A = (\text{MDF}_A - \text{MDF}_I)$ suggest that the Site M magnetites are mainly SD or PSD, with a tendency for the SD fraction to become more dominant in the high intensity zone. From J_s measurements, the total magnetite concentration in M25 averages $0.12\% \pm 0.01\%$ on a carbonate-free basis, with slightly lower abundances below 23 cm. The lack of correspondence between the NRM and bulk magnetic properties probably reflects the small fraction of magnetic carriers contributing to the NRM.

Physical remanence 'lock-in' mechanisms such as bioturbation-induced mixing or fixation due to changes in water content do not readily explain the NRM increase, because active bioturbation is limited to the topmost 5 cm where most of the change in water content also occurs. Since an Fe-enriched hydrothermal signal is difficult to distinguish from early diagenetic changes²¹, an increase in hydrothermal input cannot be completely ruled out as a source of the magnetite enrichment. However, magnetite has not been yet identified in hydrothermal effluent and any such magnetite should be associated with other hydrothermal indicators such as increased barium and amor-

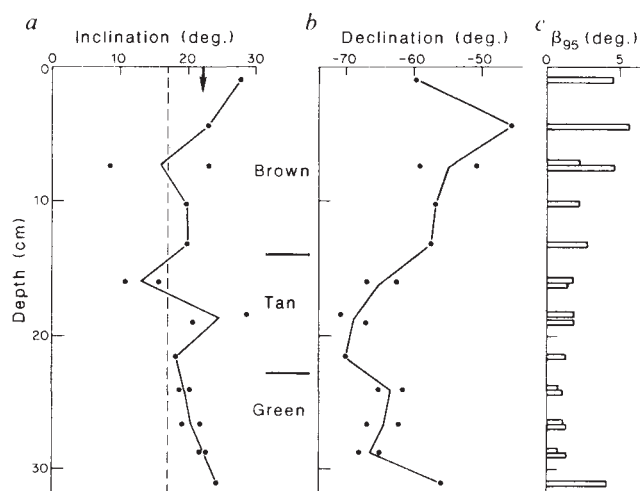


Fig. 4 a, Inclinations of M25, demagnetized at 200 Oe AF. The dashed line is the expected geocentric axial dipole inclination (17°) and the arrow is the present field direction (22°). b, Relative declinations, c, β_{95} values, based on demagnetizations at 6-7 levels from 0-400 Oe AF, showing minimum variations in the zone of high NRM intensities.

phous iron oxyhydroxides, which are not observed in M25. Furthermore, the disappearance of leachable iron oxides in the same depth range as magnetite enrichment argues for diagenesis rather than a change in source material.

The NRM and chemical results taken together, suggest that fine-grained magnetites precipitate near the top of the iron reduction zone. The same magnetic and chemical trends were found in four other cores from MANOP Site M, and we have observed similar patterns in suboxic sediments just south of the Mendicino Fracture Zone, off the Patton Escarpment, from the Weddell Sea, and from the eastern equatorial Pacific (R.K., manuscript in preparation). The rock magnetic properties of the M25 sediments are also similar to those reported by Petersen *et al.* (1986) for DSDP sediments from the South Atlantic.

In M25 and other box cores from the site, NRM intensities, MDFs and J_s decrease at the base of the core, just below the iron reduction boundary. The same behaviour was observed in P43, a 1-m gravity core taken to the west of the East Pacific Rise, except that the magnetization and MDF continue to decrease to the base of the core. As the downcore chemistry in P43 suggests no obvious source change, it appears that authigenic magnetite formation is limited to the top of the iron reduction zone.

As the entire suite of early diagenetic reactions are microbially mediated, it seems reasonable to presume that the fine-grained authigenic magnetite is formed as a result of microbial activity, for example, by magnetotactic bacteria. If so, our results indicate that magnetotactic bacterial activity is confined between the levels of nitrate reduction and the onset of more severe iron reduction. This zone is marked by the common tan-green colour transition. The microorganisms could reduce ferric oxyhydroxides and/or oxidize dissolved ferrous iron from the underlying iron reduction zone as sources of energy to oxidize organic matter, with magnetite formed as a metabolic product. The biogenic magnetites, in turn, would provide energy for iron reducers deeper in the sediment column.

In any case, the impact of authigenic magnetite on the magnetic properties of suboxic sediments can be profound. The doubling of NRM intensity, the concurrent stability increase, and other properties of M25 attributable to diagenetic magnetite have a marked effect on its magnetic stratigraphy. This chemical remanence adds a degree of complexity to palaeofield hindcasts that has yet to be unravelled.

We thank G. Frost, D. Schultz, J. Robbins and R. Chartier for laboratory assistance, and R. Merrill, H. P. Johnson and S. Levi for helpful discussions. This work was supported by NSF grant OCE-8417113.

Received 8 September 1986; accepted 23 January 1987.

- Stumm, W. & Morgan, J. J. *Aquatic Chemistry* (Wiley, New York, 1970).
- Claypool, G. & Kaplan, I. in *Natural Gases in Marine Sediments* (ed. Kaplan, I. R.) 99–139 (Wiley, New York, 1974).
- Froelich, P. N. et al. *Geochim. cosmochim. Acta* **43**, 1075–1090 (1979).
- Berner, R. A. *Early diagenesis, a Theoretical Approach* (Princeton Press, New Jersey, 1980).
- Lyle, M. *Limnol. Oceanogr.* **28**, 1026–1033 (1983).
- Hartmann, M., Müller, P. J., Suess, E. & van der Weijden, C. H. 'Meteor' *Forschungsergeb. C24*, 1–67 (1976).
- Winters, G. V. & Buckley, D. E. *Geochim. cosmochim. Acta* **50**, 277–288 (1986).
- Jahnke, R., Heggie, D., Emerson, S. & Grundmanis, V. *Earth planet. Sci. Lett.* **61**, 233–256 (1982).
- Emerson, S. in *The Carbon Cycle and Atmospheric CO₂: Natural Variations Archean to Present* (eds Sunquist, E. & Broecker, W. S.) 78–87 (AGU Geophys. Monog. 32 Washington DC, 1986).
- Blakemore, R. *Science* **190**, 377–379 (1975).
- Frankel, R. B., Blakemore, R. P. & Wolfe, R. S. *Science* **203**, 1355–1356 (1979).
- Stolz, J. F., Chang, S.-B. R. & Kirschvink, J. L. *Science* **321**, 849–851 (1986).
- Moench, T. T. & Konetzka, W. A. *Archs. Microbiol.* **119**, 203–212 (1978).
- Mann, S., Frankel, R. B. & Blakemore, R. P. *Nature* **310**, 405–407 (1984).
- Mann, S. in *Magnetite Biomineralization and Magnetoreception in Organisms* (eds Kirschvink, J. L., Jones, D. S. & MacFadden, B. J.) 311–332 (Plenum, New York, 1985).
- Kirschvink, J. L. *Earth planet. Sci. Lett.* **59**, 388–392 (1982).
- Chang, S.-B. R. & Kirschvink, J. L. in *Magnetite Biomineralization and Magnetoreception in Organisms* (eds Kirschvink, J. L., Jones, D. S. & MacFadden, B. J.) 647–669 (Plenum, New York, 1985).
- Petersen, N., von Döbeneck, T. & Vali, H. *Nature* **320**, 611–615 (1986).
- Graybeal, A. L. & Heath, G. R. *Geochim. cosmochim. Acta* **48**, 965–975 (1984).
- Lyle, M., Heath, G. R. & Robbins, J. M. *Geochim. cosmochim. Acta* **48**, 1705–1715 (1984).
- Kadko, D. *Earth planet. Sci. Lett.* **51**, 115–131 (1980).
- Klinkhammer, G. P. *Earth planet. Sci. Lett.* **49**, 81–101 (1980).
- Robbins, J. M., Lyle, M. & Heath, G. R. *A Sequential Extraction Procedure for Partitioning Elements among Co-existing Phases in Marine Sediments*, 84–93 (Oregon State Univ., 1983).
- Kent, D. V. *Science* **299**, 538–539 (1982).
- Banerjee, S. K. & Moskowitz, B. M. in *Magnetite Biomineralization and Magnetoreception in Organisms* (eds Kirschvink, J. L., Jones, D. S. & MacFadden, B. J.) 17–41 (Plenum, New York, 1985).
- Lowrie, W. & Fuller, M. J. *geophys. Res.* **76**, 6339–6349 (1971).
- Johnson, H. P., Lowrie, W. & Kent, D. V. *Geophys. J.R. astr. Soc.* **41**, 1–10 (1975).
- Levi, S. & Merrill, R. T. J. *geophys. Res.* **90**, 309–314 (1978).

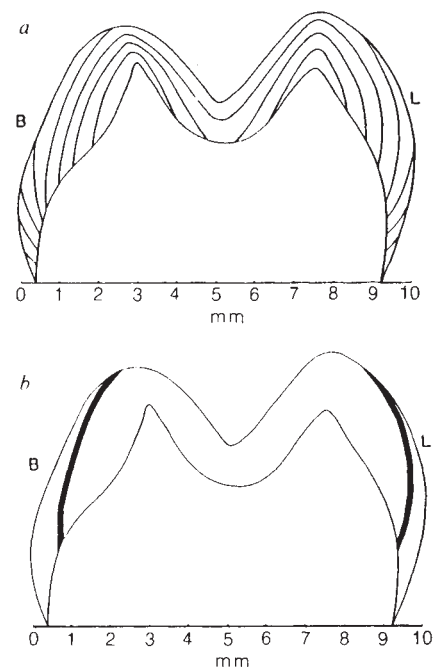


Fig. 1 Modern man: composite diagram of buccolingual sections through the mesial cusps of 10 modern M_1 teeth. *a*, Representative striae of Retzius. Striae only reach the surface on the lateral or cervical aspects of the crown, being covered by succeeding enamel increments over the cusps. *b*, The boundary between cuspal enamel in which striae do not reach the surface, and cervical (lateral) enamel where they do, is accentuated by a heavy line. B, Buccal; L, Lingual sides.

Patterns and rates of enamel growth in the molar teeth of early hominids

A. D. Beynon* & B. A. Wood†

* Department of Oral Biology, University of Newcastle-upon-Tyne, Newcastle-upon-Tyne, NE2 4BW, UK

† Department of Human Anatomy and Cell Biology, University of Liverpool, PO Box 147, Liverpool L69 3BX, UK

A recent study¹ of the surface manifestation of incremental lines associated with enamel formation suggested that the crowns of early hominid incisor teeth were formed more rapidly than those of modern humans. In the absence of comparative data, the authors were forced to assume that enamel increments in fossil teeth were similar to those in modern humans. We have used evidence from the fractured surfaces of molar teeth to deduce estimates for both long- and short-period incremental growth markers within enamel in east African 'robust' australopithecine and early *Homo* teeth. We conclude that in these early hominids, crown formation times in posterior teeth, particularly in the large thick enamelled² molar teeth of the east African 'robust' australopithecines, were shorter than those of modern humans. This evidence, considered together with data on crown and root formation times in modern apes³, suggests that the posterior teeth in these hominids both formed and erupted more rapidly than those of modern man. These results have implications for attempts to assess dental and skeletal maturity in hominids.

Incremental growth lines, striae of Retzius, have a periodicity of ~7–8 days in modern man^{4–7}. Their arrangement suggests that enamel formation can be divided into two descriptive stages

(see Fig. 1*a* and *b*). The first, or 'cuspal', stage is the secretion of full-thickness enamel over the cusp tip; in this stage striae do not reach the enamel surface. The second, or 'cervical', stage involves the asymmetric deposition of enamel on the sides of the crown, with striae reaching the surface, producing perikymata. Bromage and Dean¹ used the numbers of perikymata on the labial surface of incisor teeth in early hominids to compare the crown formation times of modern humans, early *Homo* and *Australopithecus* (*Paranthropus*) *robustus*. Their data show that crown formation was abbreviated in the fossil hominids, particularly in *A. robustus*, but these relative timings were based on the assumption that enamel formation was similar in the three taxa, for there is no information available on the patterns and rates of growth of early hominid enamel.

This paper describes an investigation of longer- and shorter-term process indicators in enamel, namely the slope and direction of incremental lines and estimates of the daily incremental rates of enamel formation, in the molar teeth of east African 'robust' australopithecines and early *Homo*. It is not at present realistic to use more than one or two teeth from each fossil taxon for sectioning to investigate the pattern of enamel growth. Instead, we have used a larger sample made up of the broken surfaces of naturally fractured teeth. The specimens were selected on two criteria: (1) the fracture plane was orientated vertically and passed close to the dentine horns, yielding surfaces equivalent to the section plane used by Martin^{8,9}; and (2) striae of Retzius were visible when specimens were immersed in absolute ethanol and viewed in polarized light². The specimens are a subsample of those used in a study on enamel thickness of the premolar and molar dentition of Plio-Pleistocene hominids from Koobi Fora, Kenya, and Olduvai Gorge, Tanzania². Specimens attributed to the formal taxon *Australopithecus* (*Paranthropus*) *boisei* included molar teeth specimens KNM-ER 801A (RM₃) and 801C (LM₃); 802E (LM₃); 816B (M); 1171C (LM₁); 1479A (LM₁); 1819 (LM₃); 3737 (RM₁) and OH 30 (RM₁). Specimens attributed to early *Homo* included material attributed to *Homo*

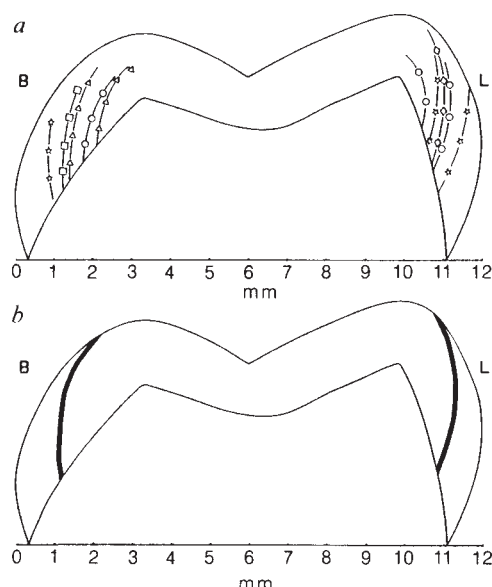


Fig. 2 Early *Homo*; crown outline constructed from sections of casts of KNM-ER 992 left and right M_1 teeth¹⁰. See text for preparation. *a*, Striae of Retzius from specimens KNM-ER 807A (☆), 809A (○), 1483E (◇), 2603 (□) and 3733 (△). *b*, The boundary between cuspal and cervical enamel interpolated as a heavy line.

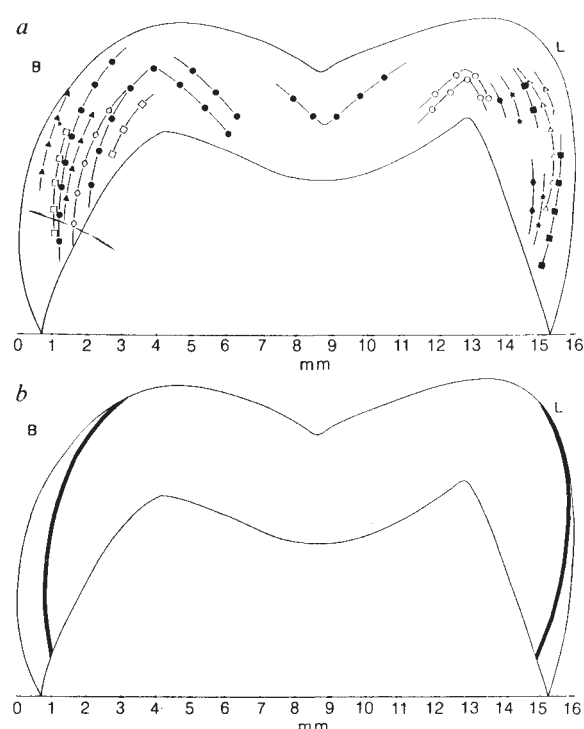


Fig. 3 'Robust' australopithecines: composite crown outline constructed from casts of KNM-ER 3230 right M_1 and M_3 (ref. 10). See text for preparation. *a*, Striae of Retzius from specimens KNM-ER 801A (◇), 801C (◆), 802E (☆), 816B (▼), 1171C (○), 1479A (●), 1819 (□), 3737 (■) and OH 30 (▽). *b*, The boundary between cuspal and cervical enamel is interpolated as a heavy line.

habilis or *Homo* sp., namely KNM-ER 807A (RM²); 809A (LM₁); 1483E (LM₂); 2603(M), and *Homo erectus* KNM-ER 3733 (LM²).

To compare the course of the striae in the two groups, the pattern of striae in each tooth was projected onto a representative tracing of a crown cross-section to build up a composite figure. Buccolingual sections through the mesial cusps of ten modern human M_1 teeth were prepared, and a composite diagram of the direction of striae of Retzius was constructed, scaled at an average buccolingual width of 10 mm (Fig. 1*a*). For the two fossil groups, reference outlines were prepared from sections made through casts of the crowns of whole teeth; the right M_1 and M_3 of KNM-ER 3230 were used as the template for *A. boisei* and the left and right M_1 of KNM-ER 992 for early *Homo*. Different templates were used because of differences in crown shape between the taxonomic categories¹⁰. Because of the progressive increase in average crown size between M_1 and M_3 in *A. boisei*¹¹, a crown outline intermediate in size between M_1 and M_3 was constructed and scaled at an average buccolingual width of 12 mm. In the early *Homo* group the M_{1-3} molar crowns are more similar in size, so the average of the profiles of the

two M_1 s of KNM-ER992 was used and scaled to a buccolingual width of 12 mm. The enamel-dentine junction (EDJ) was interpolated on each reference tracing on the basis of average enamel thickness determined at three sites on the sections², and by observing the course of the EDJ between those sites. The outlines of the enamel surface, the EDJ and the course of the striae in each tooth were traced to scale onto acetate sheets and then transferred to the composite diagrams (Figs 2*a* and 3*a*).

Inspection of the course of striae in the three groups (Figs 1*a*, 2*a* and 3*a*) suggest that the pattern of enamel deposition in robust australopithecine teeth differs from that observed in modern and early *Homo*, with most of the enamel area (90%) in the former being formed during the cuspal stage (Fig. 3*b*). This is a function of the smaller angle of slope of striae relative to the EDJ, and the thick enamel². In contrast, in both early

Table 1 Cuspal enamel completion times

Taxa	Apparent prism length (μm) l'	Cross-striation repeat interval (μm) d	Apparent formation time (d)* t'	Correction factor k	True formation time (d)* t
Modern man	1,300	5.0	260 (0.70)	1.34†	355 (0.970) ¹⁶ 340 (0.925) ¹⁷ (median 0.95)
Early <i>Homo</i>	1,600	5.8	276 (0.76)	1.34†	370 (1.01)
<i>Australopithecus</i> (<i>Paranthropus</i>) <i>boisei</i>	3,100	7.3	425 (1.16)	1.34† 1.21‡	570 (1.55) 513 (1.40)

* Time in years in parenthesis. † Modern human correction factor.

‡ Same factor reduced by 10% to compensate for relative straightness of enamel prisms².

Table 2 Cervical enamel and total crown formation times

Taxa	Cervical enamel length (μm)	Daily extension rate ($\mu\text{m d}^{-1}$)	Cervical enamel completion time (d)*	Cuspal enamel formation† (yr)	Total crown formation time (yr)
Modern man	l_c 2,000	e 3.4 ¹⁸	t_c 588 (1.60)	t 0.95	$t + t_c$ 2.55
Early <i>Homo</i>	2,000	3.4‡ 3.9§	588 (1.60)‡ 516 (1.41)§	1.01‡	2.61‡ 2.42
<i>Australopithecus</i> (<i>Paranthropus</i>) <i>boisei</i>	1,300	3.4 5.0	382 (1.04)‡ 262 (0.72)	1.55‡ 1.40¶	2.59‡ 2.12

* Time in years in parenthesis.

† From Table 1.

‡ Modern human cervical extension rate, and equivalent human completion times for cervical and cuspal enamel.

§ Corrected daily cervical extension rate (5.7:5.0).

|| Corrected daily cervical extension rate (6.3:5.0).

¶ Corrected cuspal completion time (see Table 1).

Homo (Fig. 2b) and modern humans (Fig. 1b), the area of cervical enamel is proportionally larger (25%).

A short-period indicator of incremental growth in enamel are the prism cross-striations, which have long been considered to represent daily increments of prism growth^{4,12-15}. A comparative scanning electron microscope (SEM) study on rapidly formed enamel in hominoids⁹ has given a cross-striation periodicity of 5–7 μm , with similar values (5–6 μm) in modern humans⁸; lesser values were found in the innermost and outermost enamel. There are no published data on differences in cross-striation periodicity in cuspal and cervical enamel. Observations were made on modern M_1 teeth ($n = 7$) using polarized light (PL) ($n = 6$) and an SEM ($n = 1$). There was a gradient from cuspal to cervical enamel, with mid-thickness cuspal values of $4.9 \pm 0.7 \mu\text{m}$ (PL) and $4.6 \pm 0.4 \mu\text{m}$ (SEM), reducing cervically to $3.1 \pm 0.4 \mu\text{m}$ (PL) and $3.3 \pm 0.3 \mu\text{m}$ (SEM). Preliminary SEM observations on replicas of broken hominid enamel suggest that the daily incremental rate in mid-thickness cuspal enamel in early *Homo* is $5.8 \mu\text{m d}^{-1}$, with a higher rate, $7.3 \mu\text{m d}^{-1}$, in *A. boisei*.

These observations on average daily incremental rates, combined with data on enamel thickness, allow us to make tentative estimates of crown completion times in the two fossil taxa. Prisms curve during their passage through enamel, resulting in a true length (l) greater than their apparent length (l') based on enamel thickness measurements. The true time, t , to form full thickness enamel over the cusp depends upon the cross-striation period d (which varies, being smaller at the beginning and end of the secretory cycle⁹), and the true prism length l , being greater than the apparent formation time t' estimated from enamel thickness l' divided by d . This may be expressed as $t = kt' = k(l'/d)$ where t' is the apparent formation time (l'/d) and k represents a correction factor to compensate for underestimation of l' and overestimation of d . Therefore $k = t/t' = td/l'$. The correction factor k was calculated using t values calculated from modern human M_1 teeth, and a median value for t of 0.95 years was derived from previous studies^{16,17} (Table 1). Equivalent values for t were calculated for early *Homo* and *A. boisei* (Table 1). Prisms follow a much straighter course in *A. boisei*² and two estimated values for t are presented, using either the human correction factor, and this factor reduced by 10%.

There is a substantial slowing of enamel formation during the cervical stage of crown formation in modern human molar teeth. Shellis¹⁸ has reported that the crown extension rate reduces to $3.4 \mu\text{m d}^{-1}$ in the cervical 2–3 mm of human permanent teeth. The length of cervical enamel in modern humans measured along the EDJ in Fig. 1b is approximately 2,000 μm . Assuming a daily extension rate of $3.4 \mu\text{m d}^{-1}$ (ref. 18) it would require

588 days (1.6 yr) to complete the formation of the remaining cervical enamel. Summing this figure with t (0.95 years) gives a total crown formation time for modern human M_1 teeth of 2.55 yr (Table 2). This is consistent with formation times of 2.5–3.0 yr (ref. 17) and 2.25 yr (ref. 19) reported in radiographic studies. Estimates using the modern human cervical extension rate are given for early *Homo* and *A. boisei* (1.60 and 1.04 yr). Summing these with cuspal completion times (using human correction factors) gives total crown formation times of 2.62 and 2.59 yr respectively (Table 2). It is possible that the cervical extension rate was faster in hominid enamel because the mid-thickness cross-striation interval is greater, particularly in *A. boisei*. The important value which is unknown is the interval in inner enamel, but if this value were increased by the same proportion as the differences in mid-thickness enamel this could reduce the cervical completion time in early *Homo* by 5.7:5.0, thus completing in 1.41 years, and in *A. boisei* by 7.3:5.0, giving completion in 0.72 yr. These estimates suggest a range of 2.42–2.62 yr for crown completion in early *Homo*, and 2.12–2.59 yr in *A. boisei* (Table 2). We believe that the longer times using modern human criteria are probably overestimates.

Early *Homo* molar crowns formed in a similar time to both modern man^{17,19} and modern apes³. In pongids the teeth erupt earlier with less root formed³. There is some evidence of a fast rate of root development in M_1 of early *Homo*²⁰. The conclusion, of a relatively fast rate of molar crown completion in east African 'robust' australopithecines, is consistent with perikymata counts on incisor teeth¹ of 'robust' australopithecines from southern Africa, which suggested accelerated incisor crown completion times in *A. robustus*. Despite the fact that the crowns of the molars of *A. boisei* are very much larger in overall size¹⁰, and that they have considerably thicker enamel² than modern human teeth, they apparently completed enamel formation in the same, or less, time than modern humans. This appears to be due to a combination of more rapid enamel extension and fast enamel secretion rates. The resulting pattern resembles that seen in human deciduous teeth, and may be the result of retaining in the permanent dentition the growth patterns of the deciduous teeth. Such a retention would be an example of paedomorphosis²¹, with neoteny as the particular heterochronic process²². The absolutely shorter growing period of these larger molar crowns suggests that there was strong selection to form teeth quickly and to have them erupted into the mouth early in development. The early eruption of large permanent molar teeth would have consequences for jaw growth, which would have to be rapid to accommodate them. These specializations may be yet another addition to the growing complex of characters which

is associated with increasing the size of the functional postcanine crown area in east African 'robust' australopithecines.

We thank the governments of Kenya and Tanzania, and the Director and Trustees of the National Museums of Kenya for allowing A.D.B. to examine material in their care. We are particularly grateful to P. Shellis, A. Boyde and M. C. Dean for advice and constructive comments. We thank D. J. Reid and Mrs W. Ashurst for technical assistance, and Mrs J. Rose for typing the manuscript. This research was supported by the Royal Society.

Received 23 October 1986; accepted 11 February 1987.

1. Bromage, T. G. & Dean, M. C. *Nature* **317**, 525-527 (1985).
2. Beynon, A. D. & Wood, B. A. *Am. J. phys. Anthrop.* **70**, 177-193 (1986).
3. Dean, M. C. & Wood, B. A. *Folia Primatol.* **36**, 111-127 (1981).
4. Asper, H. *Schweiz. Vjschr. Zahnheilk.* **16**, 275-314 (1916).
5. Newman, H. N. & Poole, D. F. *G. Archs. oral Biol.* **19**, 1135-1143 (1974).
6. Gustafsson, G. & Gustafsson, A. G. in *Structural and Chemical Organisation of Teeth II* (ed. Miles, A. E. W.) 75-134 (Academic, London, 1967).
7. Poole, D. F. G. in *Structural and Chemical Organisation of Teeth II* (ed. Miles, A. E. W.) 111-149 (Academic, London, 1967).
8. Martin, L. B. thesis, London Univ. (1983).
9. Martin, L. *Nature* **314**, 260-263 (1985).
10. Wood, B. A., Abbott, S. A. & Graham, S. H. *J. Anat.* **137**, 287-314 (1983).
11. Wood, B. A. & Abbott, S. A. *J. Anat.* **136**, 197-219 (1983).
12. Boyde, A. in *Scientific Foundations of Dentistry* (eds Cohen, B. & Kramer, I. R. H.) 335-352 (Heinemann, London, 1967).
13. Orban, B. *Dental Histology and Embryology* (Blakiston, Philadelphia, 1929).
14. Boyde, A. *J. dent. Res.* **58B**, 981-983 (1979).
15. Boyde, A. & Jones, S. J. *Anat. Embryol.* **168**, 211-226 (1983).
16. Massler, M. & Schour, I. *J. dent. Res.* **25**, 145-150 (1946).
17. Moorrees, C. F. A., Fanning, E. A. & Hunt, E. E. *J. dent. Res.* **42**, 1490-1502 (1963).
18. Shellis, R. P. *Archs oral Biol.* **29**, 697-705 (1984).
19. Schour, I. & Massler, M. *J. Am. dent. Ass.* **27**, 1919-1931 (1940).
20. Dean, M. C. *Am. J. phys. Anthrop.* **68**, 233-238 (1985).
21. Gould, S. J. *Ontogeny and Phylogeny* (Harvard University Press, Cambridge, Massachusetts, 1977).
22. Shea, B. T. *Science* **222**, 521-522 (1983).

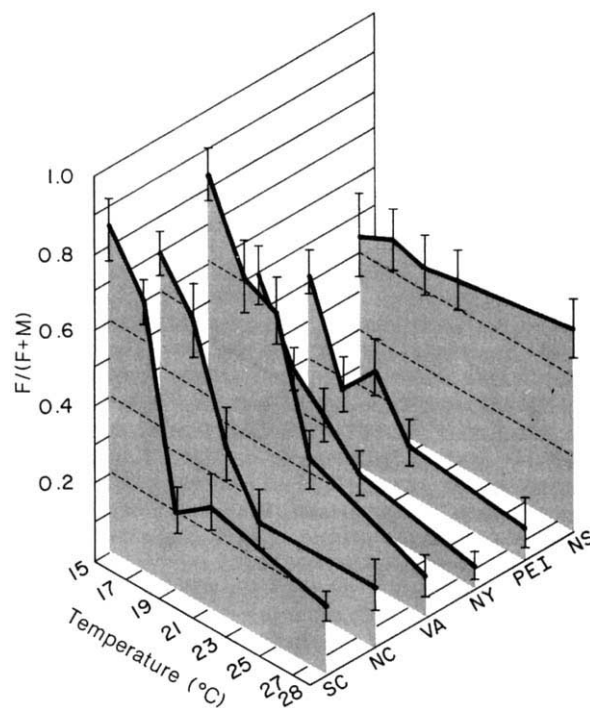


Fig. 1 Relation between sex ratio ($F/(F+M)$) and temperature during the sensitive period of sex determination in laboratory-reared progeny collected from six of the locations along the coast listed in Table 1. Locations are arranged in rank order from longest to shortest growing season. Vertical lines represent 95% binomial confidence limits.

Adaptive variation in environmental and genetic sex determination in a fish

David O. Conover & Stephen W. Heins

Marine Sciences Research Center, State University of New York, Stony Brook, 11794, USA

Two general mechanisms of sex determination have been identified among gonochoristic vertebrates: environmental sex determination where offspring become male or female in response to an environmental factor(s) during development (for example, some fishes and reptiles); and genetic sex determination where sex is determined by genotype at conception (as in birds and mammals)¹. How do these sex-determining systems evolve? Direct evidence is virtually non-existent because the sex-determining systems of most species appear to have little genetic variation²⁻⁴. Here we provide the first evidence of adaptive variation in environmental and genetic sex determination within a species. We show that in a fish with temperature-dependent sex determination, populations at different latitudes compensate for differences in thermal environment and seasonality by adjusting the response of sex ratio to temperature, and by altering the level of environmental as opposed to genetic control. The adjustments observed are precisely those predicted by adaptive sex ratio theory.

R. A. Fisher pointed out that as every offspring has one mother and one father, frequency-dependent sex ratio selection will, in many cases, lead to the evolution of genetic mechanisms that ensure a 1:1 sex ratio⁵. What conditions, however, lead to the evolution of environmental sex determination (ESD)? One view holds that ESD merely reflects phenotypic plasticity in a primitive system of sex determination⁶⁻⁸. Charnov and Bull, however, postulated that ESD is adaptive in life histories where the environment that offspring enter has a gender-dependent influence on fitness⁹. For example, if large body size is more

important to the reproductive success of females than males, those members of a population that enter an environment allowing rapid growth should become female, whereas those individuals that encounter poor growing conditions would make the best of a bad situation by becoming male.

In the Atlantic silverside, *Menidia menidia* (Pisces: Atherinidae), sex is determined by the interaction of major sex-determining genes and temperature during a distinct period of larval development^{10,11}. In nature, most offspring produced early in the breeding season, when temperatures are relatively low, become female. Most of those produced later in the breeding season, when temperatures are higher, develop into males. Females are thereby afforded a longer growing season and tend to be larger than males. Growth in the first year is the primary determinant of relative adult-body size because nearly all members (95-100%) of breeding populations are yearlings. Large size enhances at least one component of female reproductive success: a large female has a much higher fecundity than a small female, but several components of male reproductive success do not appear to be greatly affected by body size¹². Hence, an adaptive basis for ESD in *M. menidia* is plausible.

Populations of *M. menidia* inhabit salt marshes, bays and estuaries along the coast of eastern North America from northern Florida to the Magdalen Islands of Quebec, Canada¹³. Southern populations breed at somewhat higher temperatures than northern populations, and the durations of the breeding and growing seasons progressively decrease with increasing latitude¹⁴⁻¹⁶ (Table 1). Two testable predictions emerge from adaptive sex ratio theory. First, populations that experience higher overall temperatures during the breeding season compensate by shifting the functional response of sex ratio to temperature. For example, the temperature that produces a 1:1 sex ratio should be higher in southern than in northern populations. Secondly, the magnitude of ESD (that is, the maximum change in sex ratio with temperature within a population) decreases with increasing lati-

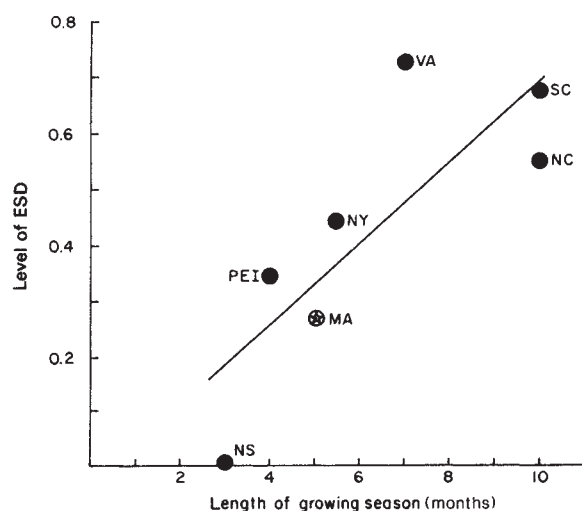


Fig. 2 Correlation between length of the growing season (months) and the level of ESD among populations along the coast. The level of ESD is defined as the proportion of fish with gender determined by temperature within each population¹¹ and estimated by the difference in sex ratio at 15 °C and 28 °C. ●, Data from Fig. 1. Star, datum derived from a previously studied Massachusetts population, representing the difference in sex ratio at 15 °C (0.43) and 21 °C (0.16), obtained by averaging across the three experiments reported in ref. 10 (no data were available at 28 °C, but there is generally little change in sex ratio between 21 and 28 °C). The equation is: $Y = -0.0312 + 0.0725X$, $r = 0.80$, $P < 0.05$, $n = 7$.

tude. This latter prediction arises because the relative benefit of being born at the beginning rather than the end of the breeding season (in terms of adult-body size) should decrease with duration of the breeding season. Moreover, inter-annual variation in mean temperature during the breeding season should be greater among the northern populations with short breeding seasons, causing sex ratio fluctuations that would select for an increased level of genetic sex determination (GSD).

Embryos of *M. menidia* were collected from six different locations along the Atlantic coast of North America (Table 1). *M. menidia* spawns *en masse* on vegetation in the intertidal zone over a 4–5 day period coinciding with new and full moons^{14,15}. We randomly removed several thousand newly fertilized eggs from many different spawning areas at each latitudinal location, 1–3 days after a semi-lunar period. The embryos were transported to the Flax Pond Marine Laboratory of SUNY Stony Brook and allowed to hatch at 21 °C. Within seven days of hatching, and before the beginning of the temperature-sensitive period¹⁷, ~100–300 larvae were randomly transferred to rearing chambers at each of five temperatures: 15, 17, 19, 21 and 28 ± 1 °C. The fish were reared at the test temperatures until reaching a size at which sex could be determined by dissection and examination of gonads¹⁷.

The temperature-sex ratio response varied greatly among populations at different latitudes (Fig. 1). Offspring from populations in South Carolina, Virginia and North Carolina had sex ratios ($F/(F+M)$) that varied from 0.70–0.85 at 15 °C to 0.10–0.18 at 28 °C. These locations correspond with the range of the southern subspecies of *M. menidia*¹³. Offspring from populations in New York and Prince Edward Island had sex ratios that ranged from 0.42–0.50 at 15 °C to 0.05–0.08 at 28 °C, and this pattern is similar to previous results from a Massachusetts population¹⁰. These are locales where the northern subspecies is found¹³. Moreover, the shape of the temperature-response curve differs with latitude. The change in sex ratio with temperature was most rapid between 17 and 21 °C among the southern populations, and in the 15–17 °C interval in the northern populations. These observations confirm our first prediction.

A clear exception to the above pattern is a population from the Annapolis Basin, Nova Scotia. Sex ratio did not differ significantly among temperatures, or from 0.5 within a particular temperature (Fig. 1). But this locale on the Bay of Fundy is itself exceptional. The Nova Scotia population inhabits an environment with the lowest summer temperatures and the shortest breeding and growing seasons of any of the populations studied^{18–21} (Table 1), even though it is at a lower latitude than Prince Edward Island. In fact, the average mid-summer sea-surface temperatures near Prince Edward Island are more similar to those of Massachusetts or New York than they are to the Bay of Fundy^{18–21}.

The absence of ESD in Nova Scotia is related to the observation that the level of ESD is highly correlated ($r = 0.80$) with length of the growing season (Fig. 2). The level of ESD was defined as the proportion of fish within each population that have gender determined by temperature, and was estimated by the difference in sex ratio between the minimum and maximum temperatures that permit sufficient growth and survival of larval *M. menidia* in the laboratory (that is, $[F/(F+M) \text{ at } 15^\circ\text{C}] - [F/(F+M) \text{ at } 28^\circ\text{C}]$)¹¹. Based on the relation in Fig. 2, the lack of ESD in Nova Scotia is not surprising: the growing season is apparently so short that the adaptive advantage of ESD has diminished.

Our second prediction has, therefore, been confirmed. As the length of the breeding and growing seasons lessen with increasing latitude, the sex ratio first becomes skewed towards the sex in which a short growing season is least harmful to reproductive success (males; for example, NY and PEI in Fig. 1). As the breeding season shortens even more, relative date of birth (early as compared to late) within the breeding season has a minimal effect on relative adult-body size, and ESD is no longer advantageous. Additionally, when the breeding season is short, temperature may become a less reliable signal of the expected length of the growing season, and mean temperature during the breeding season may vary more from year to year. Selection then leads to a genetic mechanism of sex determination with a 1:1 sex ratio independent of environment (for example NS, Fig. 1).

The transition between ESD and GSD was anticipated by

Table 1 Temperatures and length of the breeding and growing seasons among populations of *M. menidia*

Location	Latitude	Breeding season	Temperature during breeding season (°C)	Length of growing season (months)
Edisto Island, SC	32°35'	Mar–Jun	16–30	10
Beaufort, NC	34°43'	ND	ND	10
Upshurs Bay, VA	37°33'	ND	ND	7
Great South Bay, NY	40°45'	Apr–Jun	12–25	5.5
Salem, MA	42°30'	May–Jun	10–20	5
Prince Edward Island (PEI), Quebec	46°10'	ND	ND	4
Annapolis Basin, Nova Scotia (NS)	44°10'	Jun	12–22	3

Information on breeding season is from refs 14–16 (and for Great South Bay, New York from D.O.C., unpublished data). ND, No data available. Length of the growing season was defined as the number of months that near-shore sea-surface temperatures were ≥ 12 °C, based on monthly isotherm charts of the coast^{18–21}. Growth of juvenile *M. menidia* ceases at temperatures of about 12–15 °C.

population genetic models developed analytically by Bull and Bulmer^{22,23}. They showed that ESD is evolutionarily stable only when the environmental determinant of gender covaries with fitness differently for each sex, and as long as the environment does not excessively fluctuate. Otherwise, frequency-dependent sex ratio selection leads to GSD.

The Nova Scotia population also provides a unique confirmation of Fisher's sex-ratio theory. Many organisms are known to have 1:1 sex ratios but nearly all of these involve sex chromosome systems that probably lack heritable genetic variation. Hence, it could be argued that a sex ratio of 0.5 is simply a consequence of normal meiosis. A 1:1 sex ratio has evolved

under GSD in the Nova Scotia population even though variation sufficient to produce sex ratios other than 1:1 certainly exists. The adjustments in sex ratio and in the sex-determining mechanism among populations of *M. menidia* from different environments clearly illustrate the adaptive basis of ESD and GSD.

We thank L. Chiarella, M. Meade, R. Nyman and D. Monteleone for assistance in rearing the fish, D. P. Middaugh for collecting the embryos from SC, the New York Department of Environmental Conservation for providing laboratory space, and G. C. Williams and R. J. Schultz for reviewing the manuscript. This project was supported by the US NSF.

Received 15 December 1986; accepted 19 February 1987.

1. Bull, J. J. *Evolution of Sex Determining Mechanisms* (Benjamin/Cummings, London, 1983).
2. Williams, G. C. *Proc. R. Soc. B* **205**, 567-580 (1979).
3. Bull, J. J., Vogt, R. C. & Bulmer, M. G. *Evolution* **36**, 333-341 (1982).
4. Bull, J. J., Vogt, R. C. & McCoy, C. J. *Evolution* **36**, 326-332 (1982).
5. Fisher, R. A. *The Genetical Theory of Natural Selection* (Oxford University Press, 1930).
6. Kirpichnikov, V. S. *Genetic Bases of Fish Selection* (Springer, Berlin, 1981).
7. Mittwoch, U. in *Intersexuality in the Animal Kingdom* (ed. Reinboth, R.) 438-446 (Springer, Berlin, 1975).
8. Ohno, S. *Sex Chromosomes and Sex-linked Genes* (Springer, Berlin, 1967).
9. Charnov, E. L. & Bull, J. J. *Nature* **266**, 828-830 (1977).
10. Conover, D. O. & Kynard, B. E. *Science* **213**, 577-579 (1981).
11. Conover, D. O. & Heins, S. W. *Copeia* (in the press).

12. Conover, D. O. *Am. Nat.* **123**, 297-313 (1984).
13. Johnson, M. S. *Copeia* 662-691 (1975).
14. Conover, D. O. & Kynard, B. E. *Environ. Biol. Fishes* **11**, 161-171 (1984).
15. Middaugh, D. P. *Copeia* 766-776 (1981).
16. Jessop, B. M. *Can. Ms. Rept. Fish. Aquat. Sci. No. 1694* (Dept. Fisheries & Oceans, Canada, 1983).
17. Conover, D. O. & Fleischer, M. H. *Can. J. Fish. Aquat. Sci.* **43**, 514-520 (1986).
18. Walford, L. A. & Wicklund, R. I. *Serial Atlas of the Marine Environment, Folio 15* (Amersham Geographic Society, New York, 1968).
19. Colton, J. B. Jr & Stoddard, R. R. *Serial Atlas of the Marine Environment, Folio 21* (American Geographic Society, New York, 1972).
20. Lauzier, L. *Bull. Fish. Res. Bd. Can.* **111**, 251-268 (1957).
21. Lauzier, L., Trites, R. W. & Hachey, H. B. *Bull. Fish. Res. Bd. Can.* **111**, 195-212 (1957).
22. Bull, J. J. *Heredity* **46**, 9-26 (1981).
23. Bulmer, M. G. & Bull, J. J. *Evolution* **36**, 13-26 (1982).

Anti- α -fodrin inhibits secretion from permeabilized chromaffin cells

Dominique Perrin, O. Keith Langley
& Dominique Aunis*

Unité INSERM U-44, Centre de Neurochimie du CNRS,
5 rue Blaise Pascal, 67084 Strasbourg Cedex, France

Chromaffin cells release catecholamine- and peptide-containing granules by exocytosis¹⁻⁵, by a mechanism involving movement of secretory granules towards the cell membrane, their apposition to it and the fusion of the granule membrane with the plasma membrane. One of the two subunits of membrane-associated brain spectrin, α -fodrin is an actin-binding protein which is found at the periphery of chromaffin cells and may be involved in secretion⁶. Because cultured chromaffin cells can be permeabilized with detergents⁷⁻⁹, giving pores large enough to permit the entry of immunoglobulin molecules, we used permeabilized cells to test the effect of specific antibodies on secretory mechanisms. Incubation of permeabilized cells with polyclonal immunoaffinity-purified monospecific anti- α -fodrin antibody or its Fab fragments did not modify basal release but did specifically inhibit Ca^{2+} -induced catecholamine release by exocytosis. Our observations indicate that fodrin and the cytoskeleton participate in the release mechanism.

Freshly isolated or cultured chromaffin cells from bovine adrenal medulla have been successfully permeabilized with high voltage discharges¹⁰, with non-ionic detergents such as saponin⁷ or digitonin^{8,9} and recently with the α -toxin from *Staphylococcus aureus*¹¹. As membrane holes are larger in detergent-treated cells (N. J. Grant, D.A. and M. Bader, manuscript in preparation), the penetration of antibody was examined in digitonin-permeabilized cultured chromaffin cells with anti-chromogranin A antibody¹² and anti-cytochrome b_{561} antibody¹³. Chromogranin A is exclusively found in secretory vesicles¹⁴, whereas cytochrome b_{561} is a protein of the granule membrane facing the cytoplasm¹⁵. Permeabilized cells were found (Fig. 1a') to be immunoreactive with anti-cytochrome b_{561} (Fig. 1a shows immunolabelling for cytochrome b_{561} in intact cells) demonstrating that this procedure affords access of antibody to

intracellular antigenic sites of cytochrome b_{561} . A positive reaction with anti-chromogranin A on permeabilized cells would indicate that the detergents had also affected storage organelle membranes. But this was not the case (Fig. 1b, b'); chromogranin A sites remained completely inaccessible to antibody, demonstrating the integrity of secretory organelles. As shown in Fig. 1c, c', controls with anti-filamin antibody were negative indicating that antibody molecules do not bind non-specifically to digitonin-treated cells.

The effect of anti- α -fodrin, anti-vinculin and anti-filamin antibodies on the secretory activity of permeabilized cells was examined. Permeabilization and incubation with the purified immunoglobulins was in calcium-free media but in the presence of ATP¹¹, and the free calcium concentration used to trigger catecholamine release was 14.3 μM . In this system, the release rises sharply over the first 2-5 min and reaches a plateau by 10 min. Neither anti-filamin nor anti-vinculin immunoglobulins significantly affected catecholamine release, but anti- α -fodrin provoked a marked inhibition of catecholamine release. Inhibition was dependent on immunoglobulin concentration during the prestimulation period and reached 50% for a concentration of 30 $\mu\text{g ml}^{-1}$. The anti-filamin antibody does not cross-react with chromaffin cell proteins and may be considered as a control. To demonstrate the specificity of the inhibition of catecholamine release towards anti- α -fodrin, we undertook a comparative study of the effects of antibodies against other cytoskeletal proteins. Vinculin, an actin-binding protein implicated in cytoskeleton-membrane attachment^{16,17}, like fodrin, has been shown by immunofluorescence studies to be localized in the subplasmalemmal region of chromaffin cells in culture (D.P., D. Thiersé and D.A., unpublished data). Its localization is identical to that of fodrin; it forms an intensely fluorescent continuous ring at the cell periphery. But, in contrast with fodrin¹⁸, vinculin does not appear to be associated with the chromaffin granule membrane. Anti-vinculin antibody did not significantly affect calcium-induced catecholamine release from cultured chromaffin cells (Fig. 2); at 4, 10 and 20 $\mu\text{g ml}^{-1}$ catecholamine release was similar to that in cells treated with anti-filamin antibody. Thus the inhibition provoked by anti- α -fodrin antibody is specific.

To exclude a possible effect resulting from the bivalence of native immunoglobulins, Fab fragments were tested for their effect on the secretory activity of permeabilized chromaffin cells. As shown in Table 1, the incubation of detergent-permeabilized cells with increasing concentrations of anti- α -fodrin Fab frag-

* To whom correspondence should be addressed.

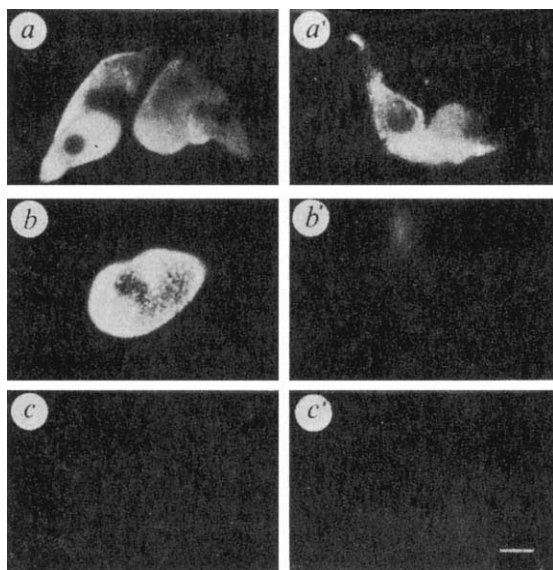


Fig. 1 Penetration of immunoglobulin molecules into digitonin-permeabilized cultured chromaffin cells. Antibodies against cytochrome b_{561} (a, a'), chromogranin A (b, b') and filamin (c, c') were used in control (a, b, c) and digitonin-treated cultures (a', b', c'). Permeabilized cells displayed intense immunofluorescence with anti-cytochrome b_{561} antibody but not with anti-chromogranin A nor with affinity-purified anti-filamin immunoglobulins.

Methods. For immunofluorescence microscopy, control cells were fixed with 4% paraformaldehyde in 25 mM sodium phosphate, 145 mM NaCl pH 7.4 (PBS), permeabilized with acetone, washed, incubated 45 min with 2 mg protein per ml normal goat immunoglobulins, then 60 min with the first antibody (anti-chromogranin A antiserum diluted 1:500, anti-cytochrome b_{561} antiserum diluted 1:150, or immunoaffinity-purified anti-filamin antibody molecules at $30 \mu\text{g ml}^{-1}$). Cells were washed three times for 15 min and antibodies were detected with swine fluorescein-conjugated anti-rabbit immunoglobulins (Nordic Immunologicals, diluted 1:50 in PBS). Preparations were examined with a Zeiss Universal Microscope using epi-illumination with 450–490 nm filters for fluorescein fluorescence⁶. Scale bar, 10 μm . For digitonin-permeabilized cells, chromaffin cells in culture were permeabilized for 10 min with 10 μM digitonin in potassium glutamate permeabilization buffer (KG buffer: 140 mM potassium glutamate, 2 mM ATP, 20 mM PIPES, 0.4 mM EGTA, 10 mM glucose, 2 mM MgSO_4 , pH 7.0) incubated 10 min in KG buffer in the presence of KG-dialysed antibody as indicated, washed briefly and fixed for 30 min with 4% paraformaldehyde in PBS, then processed for immunofluorescence microscopy with fluorescein-conjugated anti-rabbit immunoglobulin swine antiserum. Antiserum against chromogranin A was obtained after immunization of European rabbit with the antigen electrophoretically purified from soluble granule proteins¹². Antiserum against cytochrome b_{561} (ref. 13) was a gift from Dr D. K. Apps. Filamin was electrophoretically purified from chicken gizzard²⁴. The antiserum was characterized on immunoblots and purified by immunoaffinity on nitrocellulose strips. It did not cross-react with bovine filamin (D. Thiersé and D.A., unpublished data). Chromaffin cells were isolated from fresh bovine adrenal glands by retrograde perfusion with collagenase and purified on self-generating Percoll gradients^{26,27}. They were suspended at a concentration of 100,000 cells ml^{-1} in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal calf serum containing cytosine arabinoside ($3 \mu\text{g ml}^{-1}$), streptomycin ($50 \mu\text{g ml}^{-1}$) and penicillin (50 units ml^{-1}). Cells were seeded on rat tail collagen-coated glass cover-slips and cultured at 37 °C in 35-mm Petri dishes (Falcon) in a water-saturated atmosphere of 95% air + 5% CO_2 . Cells were used within 7 days of plating.

ments also resulted in inhibition of catecholamine release, reaching a value (52%) comparable to that obtained with native immunoglobulins.

To establish that this inhibitory effect is an inhibition of exocytosis, two further aspects were investigated. First, the basal release of catecholamine throughout the entire experimental procedure was followed. It was found that during the 10 min permeabilization step, the spontaneous catecholamine release

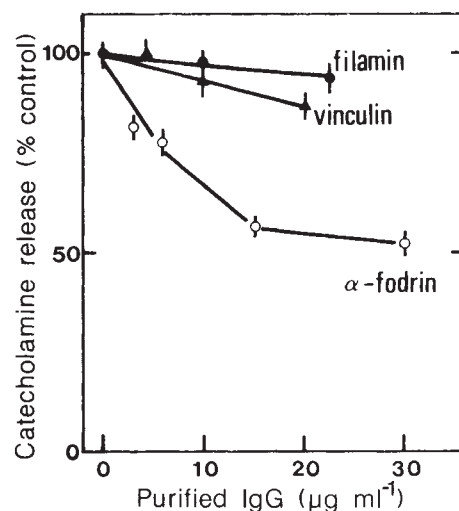


Fig. 2 Effect of specific anti- α -fodrin immunoglobulins on catecholamine release from digitonin-permeabilized cultured chromaffin cells.

Methods. Chromaffin cells were prepared as in Fig. 1, except that they were grown on 24-well Costar plates at a density of 500,000 cells per 16-mm well. Cells were loaded with [^3H] noradrenaline^{11,27} and washed for 60 min with Locke's solution (140 mM NaCl, 47 mM KCl, 1.2 mM KH_2PO_4 , 1.2 mM MgSO_4 , 11 mM glucose, 0.546 mM ascorbic acid, 15 mM HEPES, pH 7.5) containing 2.5 mM CaCl_2 . The last wash was performed with Locke's solution without calcium but containing 1 mM EGTA. Washing intervals were set constant at 10 min. Cells rinsed in calcium-free KG medium, and permeabilized with 10 μM digitonin for 10 min, then incubated with specific anti- α -fodrin, anti-vinculin or anti-filamin purified by immunoaffinity (see Fig. 1), extensively dialysed against KG medium and stimulated in KG medium containing 14.3 μM free calcium. Catecholamine release was monitored by determining the radioactivity present in the incubation medium after centrifugation for 2 min in an Eppendorf centrifuge. Cells were then precipitated with 10% trichloroacetic acid and scraped off the plates. Radioactivity was measured in scintillation vials containing 5 ml Biofluor (NEN) in an SL-4000 Intertechnique Scintillation Counter. The [^3H] noradrenaline release was calculated as per cent of total radioactivity present in the cells before stimulation. Data are expressed as per cent of noradrenaline released from control cells. Control cells release 2.6% of stored noradrenaline during the 10 min antibody incubation period and 13.8% during the 10 min stimulation period. After subtracting basal release, this net release (11.2%) was represented as 100%. Data shown are representative of three separate experiments. Each point is the mean of triplicate determinations $\pm$ s.d. Antibodies against α -fodrin⁶ were raised in rabbits with electrophoretically purified α -fodrin from bovine brain membranes²⁸. Immunoglobulin molecules were purified by immunoaffinity on nitrocellulose strips⁶. Antibodies against vinculin were raised in rabbits. Vinculin was purified from chicken gizzard²⁴ by Mg^{2+} and ammonium sulphate fractionation and chromatography on ion exchange and gel filtration columns. The highly purified preparation of vinculin was further electrophoretically purified and injected into rabbits. Immunoglobulin molecules were purified by immunoaffinity on nitrocellulose strips (D. Thiersé and D.A., unpublished). The antibody cross-reacted with bovine and rat vinculin.

amounted to $3.8 \pm 0.7\%$ ($n = 10$), similar to that during the following 10 min antibody incubation step ($2.6 \pm 0.2\%$) whichever antibody was used. Second, the possibility that intact granules were released during the stimulation period was tested. Under our conditions (10 μM digitonin, 14.3 μM free calcium with stimulation step separated from the permeabilization step), no sedimentable catecholamine was detectable in the culture medium. Therefore we conclude that anti- α -fodrin antibody acts on exocytosis.

To eliminate the possibility that the inhibition is due to a cellular redistribution of fodrin provoked by permeabilization with digitonin, the immunocytochemical localization of fodrin was compared in permeabilized and non-permeabilized cells. Using immunofluorescence or peroxidase immunohistochemistry, we showed that permeabilized chromaffin cells dis-

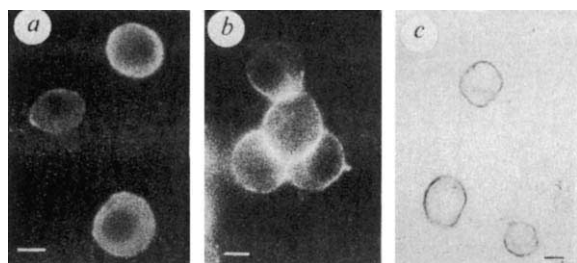


Fig. 3 Effect of digitonin on immunolabelling of chromaffin cells with anti- α -fodrin antibodies. *a*, Immunofluorescence of anti- α -fodrin in control culture permeabilized with acetone after fixation; *b*, immunofluorescence of a culture treated with digitonin before fixation; *c*, peroxidase anti-peroxidase method with no counterstaining, on a culture treated with digitonin before fixation. Bar, 10 μ m. In control cultures not permeabilized with acetone no immunolabelling is observed (not shown). Note that an identical distribution pattern for α -fodrin in the form of a peripheral ring is obtained in both control and digitonin-treated cells.

Methods. Control cells (*a*) were fixed (30 min in 4% paraformaldehyde in 0.1 M phosphate buffer, pH 7.2) and permeabilized with acetone⁶. Digitonin treatment for cultured chromaffin cells (*b*, *c*) was as described in legend for Fig. 1. Cells were then washed in KG buffer before fixation as for control cells. All cell cultures were incubated with immunoaffinity purified anti- α -fodrin antibodies⁶ followed by either (*a*, *b*) fluorescein-conjugated anti-rabbit goat antiserum (Pasteur Production, Paris, diluted 1:50) or (*c*) goat anti-rabbit IgG antiserum (diluted 1:40) and rabbit peroxidase anti-peroxidase complex (diluted 1:100; both from Sternberger-Meyer Immunocytochemicals Inc. Jarrettsville). Bound peroxidase in *c* was revealed with diaminobenzidine hydrochloride and hydrogen peroxide²⁹.

Table 1 Effect of anti- α -fodrin Fab fragments on catecholamine secretion from digitonin-permeabilized cells

Antibodies	Concentration (μ g ml ⁻¹)	Catecholamine release (%)
Non-immune IgG	40	100 $\pm$ 3
Anti-filamin IgG	50	100 $\pm$ 2
Anti- α -fodrin IgG	3	77 $\pm$ 1
	6	73 $\pm$ 1
	15	67 $\pm$ 2
Anti- α -fodrin Fab	4	83 $\pm$ 2
	10	71 $\pm$ 2
	20	52 $\pm$ 2

Chromaffin cells in culture were permeabilized as described (Fig. 1) and incubated with antibodies for 10 min, washed and stimulated with 14.3 μ M free calcium. Data are expressed as % of release from control cultures and are the means of triplicate experiments. Anti- α -fodrin Fab fragments were prepared²¹ by incubating 210 μ g of ammonium sulphate-prepared IgG with 3 μ g of papain (Worthington) in a buffer containing 100 mM sodium phosphate, 2 mM EDTA, 10 mM cysteine at pH 7.5 for 20 h at 37 °C. The solution was extensively dialysed against phosphate-buffered saline (PBS), and loaded onto a Protein A-Sepharose column (0.5 $\times$ 7.5 cm) equilibrated with PBS²². Fab fragments were eluted directly through the column; Fc fragments and intact IgGs were bound to the Protein A and were eluted with 0.2 M glycine-HCl at pH 2.8. Eluted fractions were monitored spectrophotometrically at 280 nm; protein content was calculated²³ taking $E_{280}^{1\%}$ as 13.0. Fab fragments were purified on nitrocellulose strips as described (Fig. 1 legend). Finally antibodies were dialysed against permeabilization buffer at 4 °C overnight and immediately tested.

play the same labelling pattern as intact untreated cells (Fig. 3). Electron microscopy of immunoperoxidase-labelled cultures confirmed that the peripheral ring of reaction product was continuous (Fig. 4). Frequently, as in chromaffin cells *in situ* in the adrenal gland¹⁹, secretory granules were found to be associated with this zone. We conclude that the physiological effects of anti- α -fodrin described above are not due to redistribution of fodrin in the submembrane cytoskeleton.

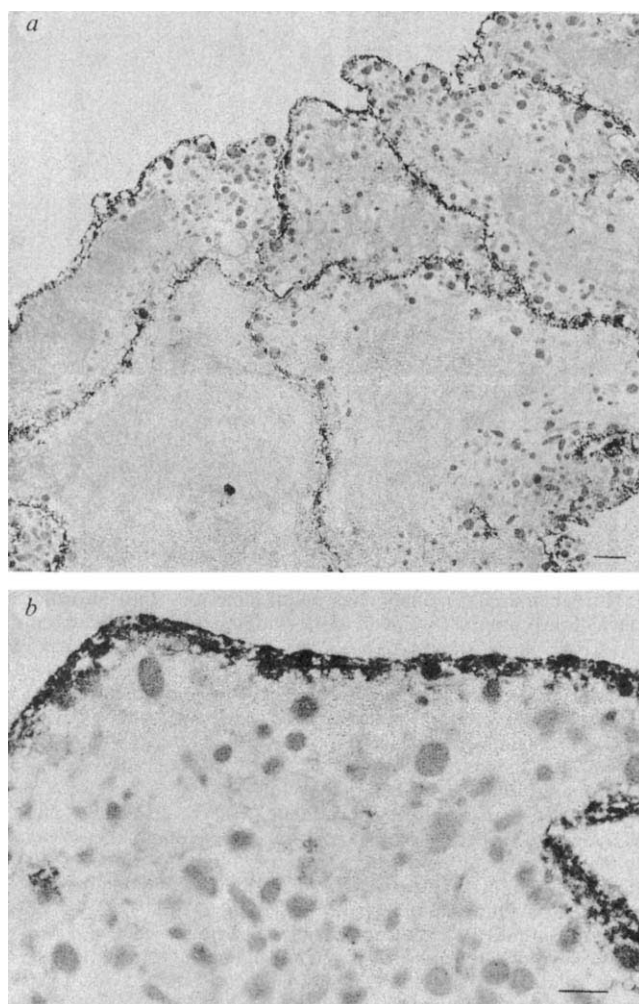


Fig. 4 Ultrastructural localization of α -fodrin in digitonin-permeabilized chromaffin cells.

Methods. Cultures immunoperoxidase-labelled for α -fodrin as outlined for Fig. 3c were treated with osmium (1 h, 1% osmium tetroxide in 0.1 M sodium cacodylate, pH 7.2), dehydrated and embedded in Spurr resin. Thin sections cut parallel to the coverslip were viewed without counterstaining in a Philips EM 420 electron microscope operated at 60 kV. Bar, 1 μ m in *a*; 0.5 μ m in *b*. Note the quasi-continuous intense labelling (*a*) of the cytoplasm close to the plasma membrane. At higher magnification (*b*) it is evident that chromaffin cell granules are frequently associated with this zone and that a filamentous substructure of the labelled material is discernible.

There are several possible reasons for the fact that the inhibition was not complete. The antibody concentration may have been too low (owing to the small amount of purified antibody available) and the incubation time too brief to permit interaction of all fodrin molecules with immunoglobulin. But Fab fragments gave a similar percentage inhibition for the same concentration, indicating that diffusion of antibody was not a limiting factor. It is also possible that antibody molecules do not react directly with the active site on fodrin concerned in exocytosis, and thus do not completely block inhibition. It may be that different granule populations exist and the population which is immediately releasable is that pre-positioned close to the plasma membrane. On stimulation these granules may be released without the control of fodrin. Other proteins, not necessarily associated with the cytoskeleton might control such release. If this is the case, it would imply that fodrin plays a role in the movement of granules and their apposition to the membrane. Our data and recent work describing rapid actin depolymerization²⁰ favour

such an hypothesis. We do not yet know what regulatory sites of fodrin are blocked by the antibody, but the data presented here directly show that fodrin and the cytoskeleton participate in the exocytotic process.

We thank Drs M. F. Bader, N. Grant and M. G. Ladona for helpful discussions, D. Thiersé, Dr D. Apps and Dr M. F. Bader for the gift of anti-vinculin and anti-filamin antibodies, anti-cytochrome *b*₅₆₁ antibody and anti-chromogranin A antibody respectively, and F. Herth for typing the manuscript.

Received 22 September 1986; accepted 4 February 1987.

- Diner, O. C. *Heb. Séanc. Acad. Sci. Paris* **265**, 616-619 (1967).
- Douglas, W. W. *Br. J. Pharmac.* **34**, 451-474 (1968).
- Smith, A. D. & Winkler, H. in *Handbk exp. Pharmac.* **33**, 538-617 (Springer, Berlin, 1972).
- Viveros, O. H. *Handbk Physiol.* **6**, 389-426 (1975).
- Burgoyne, R. D. *Biochim. biophys. Acta* **779**, 201-216 (1984).
- Perrin, D. & Aunis, D. *Nature* **315**, 589-592 (1985).
- Brooks, J. C. & Trembl, S. J. *Neurochem.* **40**, 468-473 (1983).
- Wilson, S. P. & Kirshner, N. J. *biol. Chem.* **258**, 4994-5000 (1983).
- Dunn, L. A. & Holz, R. W. *J. biol. Chem.* **258**, 4989-4993 (1983).

- Knight, D. E. & Baker, P. F. *J. Memb. Biol.* **68**, 107-140.
- Bader, M. F., Thiersé, D., Aunis, D., Ahnert-Hilger, G. & Gratzl, M. *J. biol. Chem.* **261**, 5777-5783 (1986).
- Ehrhart, M., Grube, D., Bader, M. F., Aunis, D. & Gratzl, M. *J. Histochem. Cytochem.* **34**, 1673-1682 (1986).
- Hunter, A., Waldron, K. & Apps, D. K. *FEBS Lett.* **144**, 51-56 (1982).
- Winkler, J., Apps, D. K. & Fischer-Colbrie, R. *Neuroscience* **18**, 261-290 (1986).
- Abbs, M. T. & Phillips, J. H. *Biochim. biophys. Acta* **595**, 200-221 (1980).
- Geiger, B. *Biochim. biophys. Acta* **737**, 305-341 (1983).
- Geiger, B. *Cell* **18**, 193-205 (1979).
- Aunis, D. & Perrin, D. *J. Neurochem.* **42**, 1558-1569.
- Langley, O. K., Perrin, D. & Aunis, D. *J. Histochem. Cytochem.* **34**, 517-525 (1986).
- Cheek, T. R. & Burgoyne, R. D. *FEBS Lett.* **207**, 110-114 (1986).
- Porter, R. R. *Biochem. J.* **73**, 119-126 (1959).
- Bader, M. F. & Aunis, D. *Neur. science* **8**, 165-181 (1983).
- Louvard, D., Vannier, C., Maroux, S., Pages, J. M. & Lazdunski, C. *Analyt. Biochem.* **76**, 83-96 (1976).
- Feramisio, J. R. & Burridge, K. *J. biol. Chem.* **255**, 1154-1159 (1980).
- Fenwick, C. M., Fajdiga, P. E., Howe, N. B. S. & Livett, B. G. *J. Cell Biol.* **76**, 12-30 (1978).
- Bader, M. F., Ciesielski-Treska, J., Thiersé, D., Hesketh, J. E. & Aunis, D. *J. Neurochem.* **37**, 917-933 (1981).
- Keningsberg, R. L. & Trifaro, J. M. *Neuroscience* **5**, 1547-1556 (1980).
- Davis, J. & Bennett, V. *J. biol. Chem.* **258**, 7757-7768 (1983).
- Hinton, D. M., Perrali, J. P., Meyer, H. G. & Sternberger, L. A. *J. Histochem. Cytochem.* **21**, 978-998 (1973).

Human testis-specific PGK gene lacks introns and possesses characteristics of a processed gene

John R. McCarrey*[‡] & Kelwyn Thomas[†]

* Division of Biology, Beckman Research Institute of The City of Hope, Duarte, California 91010, USA

[†] Department of Biology, California Institute of Technology, Pasadena, California 91125, USA

Phosphoglycerate kinase (PGK) (ATP:3-phospho-D-glycerate 1-phosphotransferase, EC 2.7.2.3) is a metabolic enzyme functioning in the Embden-Meyerhof pathway that converts glucose (or fructose) to pyruvate. Two functional loci for the production of PGK have been identified in the mammalian genome. *PGK-1* is an X-linked gene expressed constitutively in all somatic cells and premeiotic germ cells¹⁻⁴. The human *PGK-1* gene consists of 11 exons and 10 introns encompassing a region ~23 kilobases (kb) in length⁵. *PGK-2* is an autosomal gene expressed in a tissue-specific manner exclusively in the late stages of spermatogenesis^{3,4,6-8}. In the present study, a molecular analysis of a human genomic clone of *PGK-2* originally isolated by Szabo *et al.*⁹ has revealed that this autosomal sequence completely lacks introns and contains characteristics of a processed gene^{10,11}, or 'retropon'^{12,13}, including the remnants of a poly(A)⁺ tail and bounding direct repeats. Typically such processed sequences form non-functional pseudogenes that have evolved multiple genetic lesions which preclude translation of any transcript into a functional polypeptide¹⁰. For example, an X-linked processed pseudogene of *PGK-1* (*ψPGK-1*) in humans has been identified and shown to contain premature termination codons in all reading frames¹⁴. It was therefore unexpected to find that the intronless autosomal *PGK* sequence reported here is not a pseudogene, but is rather a functional gene that has retained a complete open reading frame, and is actively expressed in mammalian spermatogenesis. Both the unusual conservation of function in this processed *PGK-2* gene and its tissue-specific expression in spermatogenesis are best explained as a compensatory response to the inactivation of the X-linked *PGK-1* gene in spermatogenic cells before meiosis.

The nucleotide sequence determined for the autosomal genomic clone of *PGK-2* is shown in Fig. 1. An open reading frame of 1,251 nucleotides, capable of coding for a full-length PGK protein of 417 amino acids¹⁵ (predicted translation product shown) has been identified. This coding sequence shows 85% homology with that of the functional human X-linked *PGK-1*¹⁶

gene at the nucleotide level, and its predicted amino-acid sequence is 87% homologous with that of the functional *PGK-1*-derived enzyme¹⁵. The analogous locations of the 10 intervening sequences known to interrupt the *PGK-1* coding sequence¹⁴ are indicated by arrows in Fig. 1. No introns are present in this autosomal sequence, suggesting that this locus may have evolved from a spliced messenger RNA intermediate by gene processing¹⁰⁻¹³. Other indications that a processed mRNA intermediate was involved include the apparent remnant of a poly(A)⁺ tail in the 3'-untranslated region located immediately 3' to a consensus poly(A)⁺ addition signal and consisting of a tract of 165 bases of which 86 (52%) are adenines, and flanking 7-base-pair (bp) direct repeats of the sort typically added to each end of a processed transcript to facilitate its reinsertion into the genome^{10,13}. These repeats are of the sequence: 5'-TATGTTT-3', located just 3' to the poly(A)⁺ tail and at -945 to -939 5' to the AUG start codon (Fig. 1). A total of 1.4 kilobases (kb) of sequence, most of which is not shown here, has also been determined for the 5'-flanking region, including the putative promoter sequences of this gene (J.R.M., unpublished).

PGK is a highly-conserved enzyme function in all eukaryotic species, including yeast from which the PGK protein has been crystallized and subsequently characterized by structure-function studies¹⁷. The amino-acid sequence predicted by the human *PGK-2* open reading frame, along with that produced by the functional human X-linked *PGK-1* gene, and that predicted by the non-functional human X-linked *ψPGK-1* pseudogene sequence, are shown in Fig. 2, all aligned for maximum homology with the amino-acid sequence encoded by the single *PGK* gene in yeast. The predicted sequence of the *PGK-2* polypeptide is identical to the functional *PGK-1* sequence at 364/417 positions (87% homology). The 53 differences are scattered throughout the sequence with one cluster of 9 differences in 13 residues between positions 321-333. The predicted sequence of the *ψPGK-1* pseudogene protein is 91% homologous with that of *PGK-1*, indicating that the processing event which gave rise to this locus postdated that which produced the *PGK-2* gene. The yeast PGK amino-acid sequence differs at 157/417 positions from the human *PGK-2* sequence, and at 145/417 positions from the human *PGK-1* sequence, equivalent to homologies of 62% and 65% respectively. However coding information for all 17 of the amino acids found by enzyme structure-function studies to be critical for the ATP-binding site and the 'trigger mechanism' in the yeast PGK protein¹⁷ has been conserved in both the human *PGK-2* and *PGK-1* coding sequences, apparently reflecting ongoing selection at the protein level for these functional loci. By contrast, coding information for two of these same 17 amino acids has already been lost in the non-functional X-linked *ψPGK-1* pseudogene sequence.

[‡] Present address: The Division of Reproductive Biology, Department of Population Dynamics, School of Hygiene and Public Health, The Johns Hopkins University, 615 N. Wolfe Street, Baltimore, Maryland 21205, USA.

-945 TATGTTT -939
 -70 GCCCTCAAC AGCAAGTGG TTCTTCAGCA TTAAGATCCA GGTGTCAGCC TATGTCCTTA TATGTCAGG -1
 (1) ATG TCT CTT TCT AAG AAG TTG ACT TTA GAC AAA CTG GAT GTT AGA GGG AAG CGA GTC 57
 (39) Met Ser Leu Ser Lys Lys Leu Thr Leu Asp Lys Leu Asp Val Arg Gly Lys Arg Val (19)
 58 ATC ATG AGA GTA GAC TTC AAT GTT CCC ATG AAG AAG AAC CAG ATT ACA AAC AAC CAG 114
 (20) Ile Met Arg Val Asp Phe Asn Val Pro Met Lys Lys Asn Gln Ile Thr Asn Asn Gln (39)
 115 AGG ATC AAG GCT TCC ATC CCA AGC ATC AAG TAC TGC CTG GAC AAT GGA GCC AAG GCA 171
 (39) Arg Ile Lys Ala Ser Ile Pro Ser Ile Lys Lys Leu Asp Asn Gln Ala Lys Ala (57)
 172 GTA GTT CTT ATG AGT CAT CTA GGT CGG CCT GAT GGT GTT CCC ATG CCT GAC AAA TAT 228
 (58) Val Val Leu Met Ser His Leu Gly Arg Pro Asp Gly Val Pro Met Pro Asp Lys Tyr (78)
 229 TCC TTA GCA CCT GTT GCT GTT GAG CAC AAA TCC TTG CTG GGC AAG GAT GTT CTG TTC 285
 (77) Ser Leu Ala Pro Val Ala Val Glu Leu Lys Ser Leu Leu Gly Lys Asp Val Leu Phe (95)
 286 CTG AAG GAC TGT GTA GGC GCA GAA GTG GAG AAA GCC TGT GGC AAC CCA GCT CCT GGT 342
 (96) Leu Lys Asp Cys Val Glu Ala Glu Val Val Lys Lys Ala Cys Ala Asn Pro Ala Pro Gly (114)
 343 TCA GTC ATC CTG CTG GAG AAC CTG CGC TTT CAT GTG GAG GAA GAA GGG AAG GGC CAA 399
 (115) Ser Val Ile Leu Leu Glu Asn Leu Arg Phe His Val Glu Glu Glu Lys Lys Lys Gln (133)
 400 GAT CCC TCT GGA AAG AAT AAA GCT GAG CCA GAT AAA ATA GAA GCC TTC CGA GCA 456
 (134) Asp Pro Ser Gly Lys Lys Ile Lys Ala Glu Pro Asp Lys Ile Glu Ala Phe Arg Ala (152)
 457 TCA CTT TCC AAG CTA GGG GAC GTC TAT GTC AAT GAT GCT TTT GGC ACT GCA CAC GGC 513
 (153) Ser Leu Ser Lys Leu Gly Asp Val Val Lys Asn Asp Ala Phe Gly Thr Lys His Arg (171)
 514 GCT CAT AGT TCC ATG GTG GGA GTG AAT CTG CCC CAT AAA GCA TCC GGA TTC TTG ATG 570
 (172) Ala His Ser Ser Met Val Gly Val Asn Leu Pro His Lys Ala Ser Gly Phe Leu Met (190)
 571 AAG AAG GAA CTA GAT TAC TTT GCT AAA GGC TTG GAA AAC CCA GTG AGA CCC TTT CTG 627
 (191) Lys Lys Glu Leu Asp Tyr Phe Ala Lys Ala Leu Glu Asn Pro Val Arg Pro Phe Leu (209)
 628 GCT ATA CTT GGT GGA GCC AAA GTC GCA AAG ATC CAA CTT ATC AAA AAT ATG CTG 684
 (210) Ala Ile Lys Gly Gly Ala Lys Ala Gln Lys Ala Asp Lys Ile Gln Leu Ile Lys Asn Met CTG (228)
 685 GAC AAA GTC AAT GAG ATG ATT ATT GGT GGT GGA ATG GCT TAT ACC TTC CTT AAG GTA 741
 (228) Asp Lys Val Asn Glu Met Ile Thr Glu Gly Glu Asn Ala Lys Val Thr Thr Leu Lys Val (247)
 742 CTC AAC AAC ATG GAG ATT GGT GCT TCC TTT GAT GAA GAG GGA GCC AAG ATC GTT 798
 (248) Leu Asn Asn Met Glu Ile Gly Ala Lys Ser Leu Phe Asn Glu Glu Gly Ala Lys Ile Val (266)
 799 AAA GAT ATC ATG GCC AAA GCA CAA AAG AAT GGT GTA AGG ATT ACT TTT CTT GTT GAT 855
 (249) Lys Asp Ile Met Ala Lys Ala Gln Lys Asn Glu Val Arg Ile Thr Phe Pro Val Asp (286)
 856 TTT GTT ACT GGG GAC AAG TTT GAG GAC AAC GCT CAG GTT GGA AAA GCC ACT GTA GCA 912
 (286) Phe Val Thr Gly Asn Lys Phe Thr Glu Ala Lys Val Glu Lys Ala Thr Val Ala Thr (304)
 913 TCT GGC ATA TCT CTT GGC TGG ATG GGT TTG GAC TGT GGT CCT GAG AGC AAC AAG AAT 969
 (305) Ser Gly Ile Ser Pro Gly Trp Met Gly Lys Asp Cys Gly Pro Glu Ser Asn Lys Asn (323)
 970 CAT GCT CAA GTT GTG GCT CAA GCA AGC CTA ATT GTT TGG AAT GGG CGG TTA GGA GTA 1026
 (324) His Ala Glu Val Val Ala Glu Val Thr Lys Leu Ile Val Trp Asn Gly Pro Leu Gly Val (342)
 1027 TTT GAA TGG GAT GCC TTT GCT AAG GCA ACC AAA GCC CTC ATG GAT GAA ATT GTG AAA 1083
 (343) Phe Glu Thr Asn Ala Phe Ala Lys Thr Lys Thr Lys Leu Met Asp Glu Thr Val Lys (361)
 1084 GCC ACT TCC AAG GGC TGC ATC ACT GTT ATA GGG GGT GGA GAC ACT GCT ACT TGC GTT 1140
 (362) Ala Thr Ser Lys Gly Cys Ile Thr Val Ile Gly Gly Gly Asp Thr Ala Thr Cys Cys (380)
 1141 GCC AAA TGG AAC ACT GAA GAT AAA GTC AGC CAT GTC AGC ACT GGA GCG GGT GCC AGT 1197
 (381) Ala Lys Thr Asn Thr Thr Gly Asp Thr Thr Lys Thr Lys Thr Lys Thr Lys Thr Lys (399)
 1198 CTA GAG CTT CTG GAA GGT AAA ACT CTT CCT GGA GTA GAG GCC CTC AGC AAC ATG TAG 1254
 (400) Leu Glu Thr Leu Glu Gly Thr Lys Thr Lys Thr Lys Thr Lys Thr Lys Thr Lys Thr (417)
 1255 TTAATATAGT GTTACTTCT TCCTTTTCT GTCCATGCC CTAAATGAC CTAATGCTT TTACATCTCG 1324
 1325 ATGTGACCTT TGTTAAATCT TACTCTTAGA TCAAGACCTA TGTAAATGAC AAGCAGAGG CCATCAGGAA 1394
 1395 CTCTAATAT CAGCAGACGA ATTCATTITTA GTTGTGTCAC GCATTGTGCT GTTCAAGTTC TCATTGTGAC 1464
 1465 TTCCACATTT TGCTATCTAG GGAGAGCATA TTCTTGAAGT GCTATTATTA GAAAGTGAGC TGAGAAACT 1534
 1535 GAATCTTTT ATTTTATGCT AACTTGTCTA TGTGTTCTTA ATTGAAGC CAAAGATAA AACTTAATTT 1604
 1605 GTTGGGAAG GGTGAATGA AACTTGAACA ACAACAATA AATATGCCA AATAACTGA GAAATTAAT 1674
 1675 TACATATAA GAATCATGG GTACCATTA AGTTCTCTA AGGGGAATG TTATCTATA TGAAGACAA 1744
 1745 AAAAAGAGA ATGAAGACA CTATATATA CAAGAAAT TTATAAAGA AAGAAATTA ACTCAGGT 1814
 1815 AGAATATGT TGTGTGCAAT TATCAAT 1841

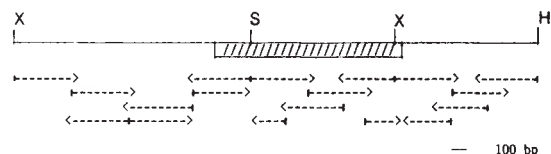


Fig. 1 *PGK-2* gene sequence. Top, restriction map and sequencing strategy. Restriction sites: X, *XbaI*; S, *SstI*; H, *HindIII*. Hatched box, *PGK* coding sequence; dotted arrows, clones sequenced. Below, nucleotide sequence with predicted amino acids. An open reading frame of 1,251 bp capable of coding for a *PGK* protein of 417 amino acids (predicted translation product shown), as well as 70 bp of 5'-untranslated sequence and 587 bp of 3'-untranslated sequence are shown. Arrows, analogous locations of the 10 introns found to interrupt the human X-linked *PGK-1* gene coding sequence¹⁴. No introns interrupt this autosomal *PGK-2* gene coding sequence. Other sequence landmarks include: underlines, direct 7-bp repeats at -945 to -939 5' to the coding sequence and at 1,820-1,826 in the 3' region and a poly(A)⁺ addition signal at 1,641-1,646; hatched underline, remnants of a poly(A)⁺ tail at 1,654-1,818 of which 86/165 nucleotides are adenines. Note that a second candidate for a poly(A)⁺ addition signal occurs at 1,655-1,660, which could have directed polyadenylation starting at 1,666. Dots over specific nucleotides, differences between the autosomal *PGK-2* sequence and the X-linked *PGK-1* nucleotide sequence.

Methods. Szabo *et al.*⁹ isolated a λ clone with an insert of 17.3 kb and an 8-kb subclone of it (pGK-3) containing sequences homologous to the *PGK-1* cDNA probe¹⁶. For our nucleotide sequence analysis, three fragments of pGK-3 were cloned into M13mp18 and M13mp19 phage³³; a 1.7 kb *XbaI-SstI* fragment that contains the 5'-flanking sequence and the first 252 bases of the coding sequence; a 0.95-kb *SstI-XbaI* fragment that contains the next 945 bases of coding sequence, and a 1.0-kb *XbaI-HindIII* fragment that contains the last 58 bases of coding sequence and the 3'-flanking sequence. The nucleotide sequence was determined using the dideoxy method³⁴. Initially a 17-base universal primer was used for each M13 clone. In each case, after 400-500 bases had been determined using this primer, a new 17-base primer, specific for the next region in the insert sequence, was synthesized and another 200-500 bases were determined. In all, 14 *PGK*-specific primers were used in addition to the universal primer. Nucleotide sequence information for the entire open reading frame as well as a majority of the flanking sequences was confirmed by sequencing both strands. DNA sequence data analysis was carried out on a Hewlett-Packard 1000 computer using DNA sequence analysis programs provided by T. Hunkapiller of Caltech³⁵.

PGK-Y PR	1	MSLSKSLVQDLQDKRVFIRVDFNVPLDGKKITSNQRIVAALPTIKYVLEHHPYVVL	60
PGK-1 PR	1	...N..TLDK...V.G...VM.....MKNNQ..N...K..V.S..FC.DNGAKS...	60
PGK-2 PR	1	...K..TLDK...VRG...IM.....MKNNQ..N...K..S.I..C.DNGAKA...	60
ψ PGK-1 PR	1	...N..TLDKP.V.GQ..IM.....VXNNQ..N...K..TVLS..FC.DNGAKS...	60
PGK-Y PR	61	ASHLGRPNGERNE-KYSLAPVAKELQSLGKDVFTLNDVGPVEAAVKASAPGSVILLE	119
PGK-1 PR	61	M.....D.VPMPD.....E..V..K.....L..K.....K.CANP.A.....	120
PGK-2 PR	61	M.....D.VPMPD.....V..K.....L..K.....K.CANP.A.....	120
ψ PGK-1 PR	61	M.....Q.D.VPMPD?.....E.F.V.FK.....L..K.....K.CANP.A.....	120
PGK-Y PR	120	NLRYHIEEGSRK-VGQKVKASKEDVQKFRHLSLADVYINDAFGTAHRAHSSVMGFD	178
PGK-1 PR	121	...F.V...KG.DAS.N...EPAKIEA..AS..K.G...V.....VN	180
PGK-2 PR	121	...F.V...KGQDPS.K.I...EPDKIEA..AS..K.G...V.....VN	180
ψ PGK-1 PR	121	...HF.V...KG...?S.N...?EPAKIEA..AS..K.GN.HV.G.....VS	179
PGK-Y PR	179	LPQRAAGFLEKELKYFGKALENTPRPLFALGGAKVADKIQILNLDKVDUSIIIGGGM	238
PGK-1 PR	181	...K.G...MK...N..A...S.E.....N.M...NEM.....	240
PGK-2 PR	181	...HK.S...MK...N..A...S.E.....N.M...NEM.....	240
ψ PGK-1 PR	180	...K.G...MK...N..V.V..S.X.....F...H..NIM...NEM...T	239
PGK-Y PR	239	AFTFKVLENTIEGDSIFDKAGAEIVPKLMEKAKAGVEVLPVDFIADAFSADANTKT	298
PGK-1 PR	241	...L..N.M...T.L..EE..K..KD..S.EKN..KIT.....VT..K.DEN.K.GQ	300
PGK-2 PR	241	...L..N.M...T.L..EE..K..KD..S.EKN..KIT.....VT..K.DEN.K.GQ	300
ψ PGK-1 PR	240	...L..N.M.T.T.L..EE..K..KD..SQ.EKN..KIT..I..VT.EK.DEN.R.GQ	299
PGK-Y PR	299	VTDKEGIPAGWQGLDNGPESRKLFAATVAKAKTIVNPPGVFEFEKFAAGTKALLDEVV	358
PGK-1 PR	301	A.VAS.....M...C...S.KY.EA.TR..Q.....V...W.A.R...M...	360
PGK-2 PR	301	A.VAS...SP...M...C...S.N.H.QV..Q.RL.....L..WDA.K...M..I.	360
ψ PGK-1 PR	300	A.VAS.....XM...C...S.KY.E..TW..Q.....V...W.A.Q...M...	359
PGK-Y PR	359	KSSAAGNTVIGGGDTATVAKKYGVTDKISHVSTGGGASLELLEKELPGVAFLEKXKX	416
PGK-1 PR	361	..ATSR.CIT.....CCA.WNTE..V.....V.....DA..NI-X	417
PGK-2 PR	361	..ATSK.CITV.....CCA.WNTE..V.....V.....I...EA..NM-X	417
ψ PGK-1 PR	360	..ATSR.CIT.....N...CCA.WNTE..R.....T.....DS.V.....DA.GNI-X	416

Fig. 2 *PGK* protein sequence comparison. The amino-acid sequences of the yeast *PGK* protein¹⁷ (PGK-Y PR), the human X-linked *PGK-1* protein¹⁵ (PGK-1 PR), the predicted amino-acid sequence of the *PGK-2* protein (PGK-2 PR), and the predicted amino-acid sequence of the human X-linked pseudogene¹⁴ (ψ PGK-1 PR) are shown arranged for maximum homology. The *PGK-2* protein sequence is 62% homologous with the yeast protein sequence, and 87% homologous with the *PGK-1* protein sequence. The human *PGK-1* protein sequence is 65% homologous with the yeast protein sequence. The ψ PGK-1 sequence is 91% homologous with the functional *PGK-1* protein sequence, and 79% homologous with the *PGK-2* protein sequence. The 17 amino acids deduced to be critical for function of the yeast *PGK* enzyme¹⁷ are boxed. Coding information for all 17 of these amino acids has been conserved in both functional human *PGK* genes, but similar information has been lost for two of these same 17 amino acids in the ψ PGK-1 sequence (indicated by *). X, Two in-frame termination codons in the ψ PGK-1 pseudogene sequence.

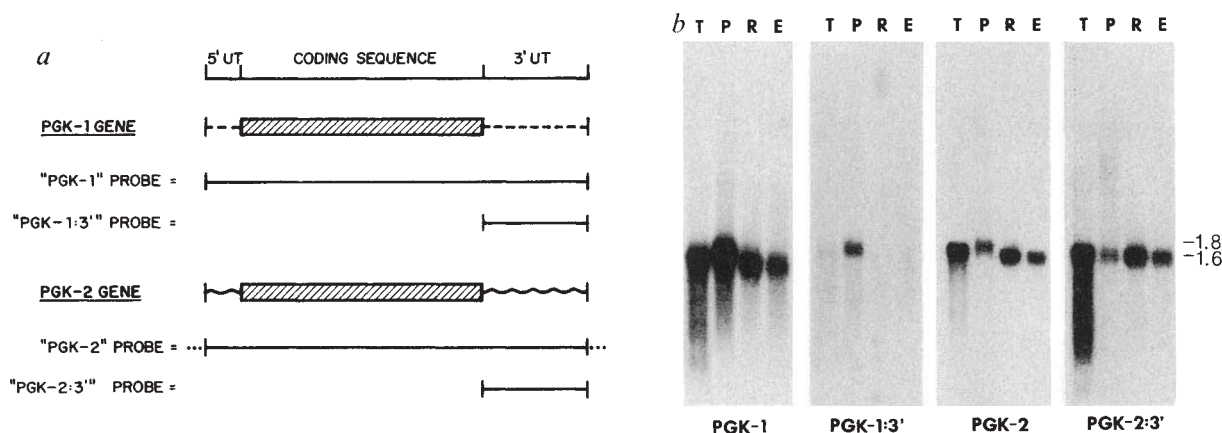


Fig. 3 Expression of *PGK* genes in spermatogenesis. *a*, Hybridization probes: four probes were used to analyse *PGK* gene expression in spermatogenesis. The 'PGK-1' probe is a full-length 1.8-kb human *PGK-1* cDNA¹⁶ that includes 80 bp of 5' untranslated sequence, 1,251 bp of coding sequence, and 437 bp of 3' untranslated sequence. The 'PGK-2' probe is the entire 3.3-kb autosomal genomic *PGK* sequence and includes 1.4 kb of 5' flanking sequence, 1,251 bp of coding sequence and 650 bp of 3' flanking sequence. These two full-length probes share a region of 85% homology in the coding sequences, and therefore cross-hybridize. Specific hybridization probes representing the 3'-untranslated region of the human *PGK-1* cDNA and the corresponding immediate 3' flanking sequence of the human autosomal *PGK* locus were prepared by appropriate restriction endonuclease digestion to generate the desired fragments: a *Bam*HI/*Sca*I double digest to produce a 434-bp fragment from the *PGK-1* cDNA (probe 'PGK-1:3') and an *Xba*I/*Kpn*I double digest to produce a 500-bp fragment from the autosomal sequence ('PGK-2:3'). The products of these digests were separated by electrophoresis in low-melting point agarose, and the desired fragments recovered by phenol extraction and ethanol precipitation. These 3' probes do not cross-hybridize. *b*, Northern blot hybridizations. The urea/LiCl method³⁶ was used to prepare total RNA from four different samples of testicular tissue: complete testis from adult mice (T), complete testis from 15-day-old prepubertal mice (P) (in which spermatogenesis has only proceeded through the spermatocyte stage), and specific postmeiotic spermatogenic cell types isolated by staput fractionation³⁷, including round spermatids (R) and elongated spermatids (E). For each sample poly(A)⁺ RNA was isolated from total RNA by one passage over an oligo(dT)-cellulose column and then electrophoresed in 1% agarose containing 2.2 M formaldehyde³⁸, and blotted to GeneScreen membranes using manufacturer's conditions (NEN). The *PGK-1*, *PGK-2*, *PGK-1:3'*, and *PGK-2:3'* probes were labelled with ³²P by nick translation³⁸ to specific activities of 1–5 × 10⁸ c.p.m. μg⁻¹, and each hybridized to a separate blot containing all four testis RNA samples according to the method of Thomas³⁹. After 12–18 h of hybridization, the blots were washed in 1 × SSC (150 mM NaCl, 15 mM Na citrate), 0.1% SDS at 68 °C and exposed to X-ray film. Both full-length probes (*PGK-1* and *PGK-2*) containing cross-hybridizing coding sequences hybridized to all four RNA samples. In RNA from the total adult testis and total prepubertal testis, two transcripts, one at 1.8 kb and one at 1.6 kb, were detected, but only the 1.6 kb transcript was represented in the isolated round and elongated spermatids. The unique 3' probes produced hybridization results that defined the specific expression patterns of each *PGK* gene in spermatogenesis. Thus the *PGK-1:3'* probe detected only the 1.8-kb *PGK-1* RNA that represents a minor portion of the *PGK* RNA in total adult testis, but is the predominant component of it in prepubertal testis. The predominance of the *PGK-1* transcript at this stage is best demonstrated by the significantly stronger hybridization of both full-length probes to the 1.8-kb transcript in the prepubertal testis, as compared to the relatively weaker signal from the 1.6-kb transcript at this stage. In contrast the *PGK-2:3'* probe detected only the 1.6-kb *PGK-2* RNA that represents a majority of the *PGK* RNA seen in total adult testis, a minority of the *PGK* RNA in 15-day prepubertal testis, and the only *PGK* RNA present in postmeiotic spermatogenic cells.

To test for *in vivo* expression of the intronless autosomal *PGK* gene, we compared the hybridization of this genomic sequence with that of the *PGK-1* cDNA sequence¹⁶ to RNAs isolated from complete adult mouse testis, complete pre-pubertal mouse testis, and two isolated stages of postmeiotic spermatogenic cells, round and elongated spermatids, respectively. The results are shown in Fig. 3. Because these full-length sequences share a region of 85% homology (the open reading frame) which spans 1,251 nt (see Fig. 3a), they both cross-hybridized to all four RNA samples (*PGK-1* and *PGK-2* in Fig. 3b). However, when probes were used that represented only the non-homologous 3'-untranslated regions of each gene (see Fig. 3a), we found that *PGK-1* expression in the testis is limited to somatic cells and premeiotic germ cells (*PGK-1:3'* in Fig. 3b), whereas the *PGK-2* sequence is expressed primarily in postmeiotic spermatogenic cells (*PGK-2:3'* in Fig. 3b).

To confirm that hybridization of the *PGK-2*-specific sequence to spermatogenic cell RNAs is indeed indicative of *in vivo* transcription from the intronless *PGK-2* locus, and not an artefact of cross-hybridization to transcripts from another autosomal *PGK* locus, we analysed the organization of *PGK* sequences in the human genome using our specific 3' probes (Fig. 4). In previous work, it had been shown that hybridization of a full-length *PGK* probe to human DNA digested with *Bam*HI yields at least four X-linked *PGK* bands and three autosomal bands^{16,18}. The autosomal bands include one at 14 kb, one at 2.1 kb, and one at 1.0 kb. Of these, our *PGK-2*-specific probe (*PGK-2:3'*) hybridized uniquely to the 14-kb band (lane

c in Fig. 4). By contrast, the *PGK-1*-specific probe (*PGK-1:3'*) hybridized only to the 2.1-kb band among the autosomal sequences (lane b in Fig. 4). Neither 3' probe hybridized to the 1.0-kb autosomal *PGK* band. These results are consistent with Southern blot data from several groups^{14,18–20} supporting the existence in the human genome of two full-length autosomal *PGK* sequences, one functional (the 14-kb band) and one non-functional (the 2.1-kb band), and possibly a third, incomplete autosomal *PGK* sequence (the 1.0-kb band). Our demonstration that the 3'-untranslated sequence from the intronless autosomal *PGK* gene is both unique to a single *PGK* locus in the human genome (the 14-kb band), and hybridizes to postmeiotic spermatogenic cell RNA, indicates this is the functional *PGK-2* gene. This conclusion is further supported by the elimination of the only other full-length autosomal *PGK* sequence in the human genome (the 2.0-kb band) as a candidate for the functional *PGK-2* gene on the basis that it cross-hybridized with the *PGK-1:3'* probe, thus confirming that this locus is not expressed in the late stages of spermatogenesis, and therefore cannot be the functional *PGK-2* gene.

The autosomal genomic *PGK* sequence analysed in the present study contains all of the hallmarks of a processed gene^{10,11}, or retroposon^{12,13}, including a lack of all 10 intervening sequences found in the functional *PGK-1* gene, the remnants of a poly(A)⁺ tract in the 3'-untranslated region immediately following a poly(A)⁺ addition signal, and direct repeats 7 bp long located 5' and 3' to the gene. These features are consistent with the notion that this locus arose by reverse transcriptase-mediated

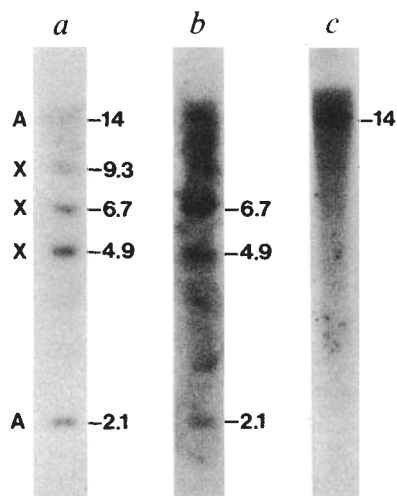


Fig. 4 Genomic organization of human PGK genes. DNA for Southern blot hybridization was prepared from human cell lines maintained in tissue culture, digested to completion with *Bam*HI restriction endonuclease, electrophoresed in 0.8% agarose, and blotted to nitrocellulose³⁸. Separate blots were probed under standard conditions³⁸ with the ³²P-labelled PGK-1 cDNA (a), the PGK-1:3' fragment (b), and the PGK-2:3' fragment (lane c), respectively. Three autosomal (14-, 2.1- and 1.0-kb) and four X-linked (9.3-, 6.7-, 4.9- and 3.8-kb) PGK *Bam*HI bands have been previously demonstrated in human genomic DNA^{16,18}. In lane a, hybridization with the PGK-1 probe detected two of the autosomal (A) bands at 14 and 2.1 kb respectively, and three of the X-linked (X) bands at 9.3, 6.7 and 4.9 kb respectively. Of these, the PGK-1:3' probe hybridized only to the 2.1-kb autosomal band and to the 6.7 and 4.9 kb X-linked bands as shown in lane b; whereas the PGK-2:3' probe hybridized uniquely to the 14-kb autosomal band as shown in lane c.

processing of a tailored RNA transcript originally produced by a separate progenitor PGK locus extant in the genome, presumably the intron-containing, X-linked PGK-1 gene.

The appearance of the PGK-2 locus predates the split between eutherian mammals and marsupials, both of which possess separate, functional X-linked PGK-1 genes and autosomal PGK-2 genes, but apparently postdates the earlier split between therian and monotreme mammals, as the latter have only a single functional PGK locus²¹. That the emergence of the PGK-2 locus was not a recent event in mammalian evolution is reflected in the 15% divergence between the coding sequences of the human PGK-1 and PGK-2 genes, and the >65% divergence in their respective 5'- and 3'-untranslated regions. Thus the apparent degeneration of the poly(A)⁺ tail remnant in the processed PGK-2 gene from a presumed original status of 100% adenines to its current status of 52% adenines is a predictable consequence of evolutionary divergence. However, for this same reason, the identification of the flanking direct repeats is not straightforward, as these sequences should also have undergone significant divergence. The suggestion that the heptamer we have identified represents the actual direct repeat is strengthened by its 3' position in the region immediately downstream of the poly(A)⁺ tail remnant, which should represent the 3' border of the original processed sequence. It is also interesting to note that this sequence is exactly homologous to a portion of the 11-bp direct repeats of the sequence 5'-ACTTATGTTTC-3' that bound the human ψ 1 β -actin processed pseudogene^{22,23}.

There have been other reports of specific, functional, intronless genes that may have evolved from intron-containing progenitor genes, including the globin genes in *Chironomus*²⁴, and a tissue-specific calmodulin gene in the chicken²⁵. However both the chicken calmodulin gene and the globin genes in *Chironomus* lack bounding direct repeats and any remnant of a poly(A)⁺ tail in their 3'-flanking regions, indicating these loci

may have eliminated introns directly by gene conversion rather than through gene processing. By contrast the rat preproinsulin I gene²⁶ does have a short poly(A)⁺ tail remnant (ACCAAAAAAAAA) at the beginning of its 3'-flanking region, and is bounded by direct repeats, indicating gene processing has occurred. However this locus represents a 'semiprocessed'²⁶ gene as only one of the two introns present in its presumed progenitor, the preproinsulin II gene, has been eliminated, while the other remains. In addition, this sequence is flanked by unusually long (41-bp) direct repeats, which are separated from the processed sequence by a span of 42 bp of unknown origin in the 5' end. Thus the human PGK-2 gene described here is the first case of a completely intronless, functional gene that demonstrates all of the predicted signs of having evolved from an intron-containing progenitor by gene processing.

The pattern of PGK gene expression detected by our 3' probes at the RNA level in spermatogenesis is in agreement with previous reports based on direct analysis at the protein level^{27,28}, and indirect analysis at the RNA level^{29,30}. Our results provide the first direct evidence at the RNA level in support of the idea that the single X-chromosome in spermatogenic cells becomes inactivated before meiosis, as originally proposed by Lifschytz and Lindsley³¹. These results also support the occurrence of transcription of the PGK-2 gene in postmeiotic, haploid spermatids, but to obtain unequivocal evidence for this phenomenon analysis of nuclear run-off transcripts with our specific 3' probes is required. These experiments are in progress.

Because so many processed pseudogenes have been described in mammals (for reviews see refs 10, 11, 13), the occurrence of a functional processed gene represents a rare finding, and suggests that relatively unusual circumstances may have predisposed its evolution. We would propose that inasmuch as mature spermatozoa require significant amounts of phosphoglycerate kinase to participate in the metabolism of fructose present in semen and secreted in both the male and female reproductive tracts³², the inactivation of the single X-chromosome in spermatogenic cells before meiosis may have contributed to a need for a functional autosomal PGK locus. Thus in this case, the random event of gene processing could have given rise to the requisite autosomal PGK gene, which was subsequently conserved as a functional locus to fulfil this need.

We thank P. Szabo and M. Siniscalco for making their genomic PGK clone available; V. Coghlan, J. Kaufman, G. Morgan and A. Siu for technical assistance; and A. Riggs, S. Ohno and M. Simon for valuable discussions. This work was supported by a grant from the NIH (HD 18360) to J.R.M. and a grant from the American Cancer Society (CD-294). The sequence data in this paper have been submitted to the EMBL/GenBank Data Libraries under the accession number Y00261.

Received 8 July 1986; accepted 16 January 1987.

- Chen, S.-H., Malcolm, L. A., Yoshida, A. & Gilbert, E. R. *Am. J. hum. Genet.* **23**, 87-91 (1971).
- Kozak, L. A., McLean, G. K. & Eicher, E. M. *Biochem. Genet.* **11**, 41-47 (1973).
- VandeBerg, J. L., Cooper, D. W. & Close, P. J. *J. exp. Zool.* **198**, 231-240 (1977).
- Kramer, J. M. & Erickson, R. P. *Dev. Biol.* **87**, 37-45 (1981).
- Michelson, A. M., Blake, C. C. F., Evans, S. T. & Orkin, S. H. *Proc. natn. Acad. Sci. U.S.A.* **82**, 6965-6969 (1985).
- Eicher, E. M., Cherry, M. & Flaherty, L. *Molec. gen. Genet.* **158**, 225-228 (1978).
- VandeBerg, J. L., Cooper, D. W. & Close, P. J. *Nature new Biol.* **243**, 48-50 (1973).
- VandeBerg, J. L., Cooper, D. W., Sharmar, G. B. & Poole, W. E. *Genetics* **95**, 413-424 (1980).
- Szabo, P., Greshchik, K.-H. & Siniscalco, M. *Proc. natn. Acad. Sci. U.S.A.* **81**, 3167-3169 (1984).
- Vanin, E. F. *A. Rev. Genet.* **19**, 253-272 (1985).
- Wagner, M. *Trends Genet.* **2**, 134-137 (1986).
- Rogers, J. H. *Nature* **301**, 460 (1983).
- Rogers, J. H. *Int. Rev. Cytol.* **93**, 187-279 (1985).
- Michelson, A. M., Bruns, G. P., Morton, C. C. & Orkin, S. H. *J. biol. Chem.* **260**, 6982-6992 (1985).
- Huang, L.-Y., Welch, C. D. & Yoshida, A. *J. biol. Chem.* **255**, 6412-6420 (1980).
- Singer-Sam, J. *et al. Proc. natn. Acad. Sci. U.S.A.* **80**, 802-806 (1983).
- Watson, H. C. *et al. EMBO J.* **1**, 1635-1640 (1982).
- Singer-Sam, J. *et al. Gene* **32**, 409-417 (1984).
- Willard, H. F., Goss, S. J., Holmes, M. T. & Munroe, D. L. *Hum. Genet.* **71**, 138-143 (1985).
- Gartler, S. M. *et al. Somatic molec. Cell Genet.* **12**, 395-401 (1986).
- VandeBerg, J. L., Cooper, D. W., Sharmar, G. B. & Poole, W. E. *Genetics* **95**, 413-424 (1980).
- Moos, M. & Gallwitz, D. *Nucleic Acids Res.* **10**, 7843-7849 (1982).
- Moos, M. & Gallwitz, D. *EMBO J.* **2**, 757-761 (1983).
- Antoine, M. & Niessing, J. *Nature* **310**, 795-798 (1984).

25. Stein, J. P. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **80**, 6485-6489 (1983).
26. Soares, M. B. *et al.* *Molec. cell. Biol.* **5**, 2090-2103 (1985).
27. Kramer, J. M. *Dev. Biol.* **87**, 30-36 (1981).
28. VandeBerg, J. L., Lee, C.-Y. & Goldberg, E. J. *exp. Zool.* **217**, 435-441 (1981).
29. Erickson, R. P., Kramer, J. M., Rittenhouse, J. & Salkeld, A. *Proc. natn. Acad. Sci. U.S.A.* **77**, 6086-6090 (1980).
30. Gold, B., Fujimoto, H., Kramer, J. M., Erickson, R. P. & Hecht, N. B. *Dev. Biol.* **98**, 392-399 (1983).
31. Lifschitz, E. & Lindsley, D. L. *Proc. natn. Acad. Sci. U.S.A.* **69**, 182-186 (1972).
32. Mann, T. *The Biochemistry of Semen and the Male Reproductive Tract* (Wiley, New York, 1964).
33. Yanisch-Perron, C., Vieira, J. & Messing, J. *Gene* **33**, 103-119 (1985).
34. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
35. Hunkapiller, M. *et al.* *Nature* **310**, 105-111 (1984).
36. Drubin, D. G., Caput, D. & Kirschner, M. W. *J. Cell Biol.* **98**, 1090-1097 (1984).
37. Romrell, L. J., Bellvé, A. R. & Fawcett, D. W. *Dev. Biol.* **49**, 119-131 (1976).
38. Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning, a Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).
39. Thomas, P. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5201-5205 (1980).

Thy-1-mediated T-cell activation requires co-expression of CD3/Ti complex

Kurt C. Gunter, Ronald N. Germain,
Richard A. Kroczeck, Takashi Saito, Wayne M. Yokoyama,
Christina Chan, Arthur Weiss* & Ethan M. Shevach

Laboratory of Immunology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA

* Howard Hughes Medical Institute and the Department of Medicine, University of California San Francisco, San Francisco, California 94143, USA

In addition to monoclonal antibodies against the CD3 (T3)-T-cell antigen receptor (CD3/Ti) complex¹⁻³, several other monoclonals directed towards distinct cell surface structures on human (CD2 (T11)⁴ and Tp44^{5,6}) and murine (Thy-1⁷⁻⁹, TAP¹⁰, and Ly-6¹¹) T lymphocytes are capable of activating T cells. It has been proposed that such structures may function as alternative pathways of stimulation. To examine directly whether any relationship exists between Thy-1-dependent activation phenomena and T-cell activation mediated through the CD3/Ti complex, we have transfected several CD3/Ti⁻ variants^{12,13} of the human T-cell line Jurkat with the murine *Thy-1.2* gene. Our data indicate that in CD3/Ti⁻, Thy-1.2⁺ transfectants, monoclonal antibodies against Thy-1.2 can induce a rise in cytoplasmic free calcium ([Ca²⁺]_i), but fail to stimulate interleukin-2 (IL-2) production. The only defect in these variant cell lines responsible for the inability to produce IL-2 in response to Thy-1 stimulation was in the expression of the CD3/Ti complex, because replacement of defective Ti α - or β -chain genes reconstituted both surface expression of CD3/Ti and responsiveness to Thy-1 in the IL-2 production assay.

We have previously demonstrated the ability of anti-Thy-1 monoclonal antibodies (mAbs) to elicit IL-2 secretion from Jurkat cells expressing a transfected Thy-1.2 molecule on the cell surface¹⁴. In the present experiment, two independently derived CD3/Ti⁻ Jurkat human T-cell lines (J.RT3-T3.5 and J.RT-T3.1) were selected as recipients of the murine *Thy-1.2* gene, and a genomic clone of *Thy-1.2* in the pSV2gpt vector was introduced into each. Several independent colonies grew from each transfection and expressed membrane Thy-1 as judged by flow microfluorometry (FMF). Following cloning by limiting dilution, one representative clone from each transfection was studied in greater detail (Fig. 1). For comparison, we have included the profiles obtained with J14C9, a previously characterized¹⁴ CD3/Ti⁺, Thy-1.2⁺, transfectant of the parental Jurkat line. J14C9 expresses both CD3 and high levels of Thy-1.2 (Fig. 1, *a*, *d*); in contrast, clone 18B3 (Fig. 1*b*, *e*), the Thy-1.2⁺ transfectant of J.RT-T3.1 and clone 1B3 (Fig. 1*c*, *f*), the Thy-1.2⁺ transfectant of the β -chain mutant, J.RT3-T3.5 both fail to express CD3, but stain very brightly for Thy-1.2.

In the representative experiment shown in Table 1 (top), the J14C9 (CD3/Ti⁺, Thy-1.2⁺), 18B3 (CD3/Ti⁻, Thy-1.2⁺) and

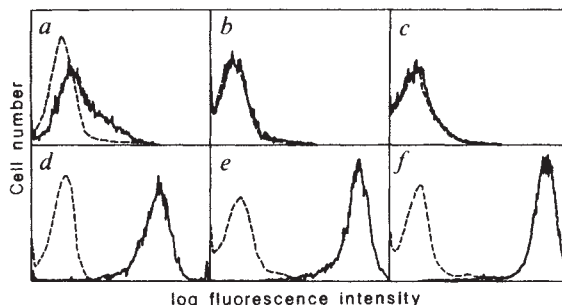


Fig. 1 OKT3 and anti-Thy-1.2 staining of CD3/Ti⁺, Thy-1.2⁺ and CD3/Ti⁻, Thy-1.2⁺ Jurkat transfectants. The CD3/Ti⁺, Thy-1.2⁺ Jurkat transfectant J14C9 (*a* and *d*) and the CD3/Ti⁻, Thy-1.2⁺ transfectants 18B3 (*b* and *e*) and 1B3 (*c* and *f*) were stained with either OKT3 ascites (*a-c*) or an anti-Thy-1.2 ascites (*d-f*). Fluorescein-isothiocyanate-(FITC)-conjugated second antibody in the absence of first antibody is shown as a control (broken line).

Methods. The general methods have been described in detail previously¹⁴. To transfect the CD3/Ti⁻ Jurkat cell lines J.RT3-T3.5 and J.RT-T3.1 with Thy-1.2, *Escherichia coli* strain HB101 bacteria containing the amplified plasmid pSV2gpt-Thy-1.2 were treated with lysozyme to produce spheroplasts that were fused to the target cell lines by standard polyethylene glycol methods. After ~48 h selection medium was added. For the genomic Thy-1.2 construct, the selection medium consisted of 1.0 μ g ml⁻¹ mycophenolic acid, 15 μ g ml⁻¹ hypoxanthine and 250 μ g ml⁻¹ xanthine. After 3-5 weeks of culture, growing colonies were analysed for Thy-1 expression by FMF. Colonies which exhibited positive staining were then cloned at limiting dilution. For FMF analysis, the cell line (10⁶ cells) selected for analysis was incubated with an excess of a mixture of rat anti-Thy-1 mAbs¹⁴ or with the anti-CD3 mAb, OKT3 for 30 min at 4°C. Cells were then washed 3 times with Hank's Balanced Salts Solution containing 0.1% Na₂S₂O₈ and 3% FCS. Cells were then reacted with either a FITC-conjugated mouse mAb specific for rat κ chains, MAR 18.5 (anti-Thy-1 staining) or a FITC-conjugated goat anti-mouse reagent (CD3 staining). After another 30 min incubation at 4°C, cells were again washed, and evaluated on a fluorescence-activated cell sorting (FACS) analyser (Becton Dickinson).

1B3 (CD3/Ti⁻, Thy-1.2⁺) cell lines were compared for their ability to secrete IL-2 in response to a number of stimuli. Previous studies with Jurkat have established the requirement for TPA (12-*O*-tetradecanoyl-phorbol-13-acetate) in conjunction with a second stimulus to induce IL-2 secretion optimally¹². In the presence of TPA, the OKT3 mAb (directed against CD3) induced IL-2 production only in the CD3/Ti⁺ J14C9 cell line. Anti-Thy-1 mAbs, either present continuously during the culture or bound to the cell surface and then cross-linked with anti-Ig antibodies, stimulated IL-2 production by J14C9, but repeatedly failed to activate 18B3 or 1B3. Although both 18B3 and 1B3 failed to produce IL-2 in response to OKT3 or anti-Thy-1, both lines were capable of producing IL-2 when stimulated by the calcium ionophore, A23187.

IL-2 secretion is a late measure of T-cell activation and it is therefore possible that inability of a cell to respond to anti-Thy-1 could reside at any of several steps of the T-cell activation cascade. We therefore examined whether the inability of the CD3/Ti⁻, Thy-1.2⁺ transfectants to respond to anti-Thy-1 extended to other parameters of the T-cell activation process. We have previously shown that following cross-linking by anti-Ig, mAbs to Thy-1 can elicit rapid rises in [Ca²⁺]_i in normal T cells, T-cell hybridomas, Thy-1.2-transfected murine B lymphoma cells¹⁵, and Thy-1.2-transfected¹⁴ Jurkat cells. Under these conditions, anti-Thy-1 mAbs were also capable of eliciting a rise in [Ca²⁺]_i in the CD3/Ti⁻, Thy-1.2⁺ transfectants (Fig. 2, 18B3 in *b*, 1B3 in *c*). The rise in [Ca²⁺]_i was in most studies indistinguishable in magnitude from the induced in the CD3/Ti⁺, Thy-1.2⁺ transfectant (J14C9, Fig. 2*a*).

These data indicated that anti-Thy-1 mAbs caused equivalent changes in [Ca²⁺]_i in CD3/Ti⁺, Thy-1.2⁺ and CD3/Ti⁻, Thy-1.2⁺

Fig. 2 Anti-Thy-1 induced changes in $[Ca^{2+}]_i$ in CD3/Ti⁺, Thy-1.2⁺ and CD3/Ti⁻, Thy-1.2⁺ transfected Jurkat cell lines. J14C9 (CD3/Ti⁺, Thy-1.2⁺, panel a), 18B3 (CD3/Ti⁻, Thy-1.2⁺, panel b) and 1B3 (CD3/Ti⁻, Thy-1.2⁺, panel c) cell lines were pre-incubated with anti-Thy-1 mAbs, then loaded with Quin 2-AM which is trapped intracellularly and shows increased fluorescence on binding Ca^{2+} . After a baseline had been established, antibody MAR 18.5 (1/100 ascites) was added (arrow). Approximate values for $[Ca^{2+}]_i$ are given on the vertical axis. Absolute values for maximum increases in $[Ca^{2+}]_i$ were: J14C9, 60 nM; 18B3, 259 nM; 1B3, 195 nM.

Methods. Cells were first preincubated with the rat anti-Thy-1 mAbs, 3A7 and 3H11¹⁴ at 1/100 ascites dilution for 20 min at 4 °C. After washing, cells were loaded at 5×10^6 ml⁻¹ in DMEM containing 10% FCS with 10 μ M Quin 2-AM (Sigma) for 30 min. The cells were then washed two times with Dulbecco's phosphate buffered saline (PBS), resuspended in the same buffer to a concentration of 5×10^6 ml⁻¹, and used immediately after equilibration at 37 °C for 5 min. The cross-linking antibody, MAR 18.5 was then added at a final concentration 1/100 ascites. Fluorescence intensity was measured with a fluorescence spectrophotometer (model LS-5; Perkin-Elmer) (excitation 399 nm, emission 492 nm). The cuvette chamber was maintained at 37 °C, and the cell suspension was continuously stirred. Maximum fluorescence (F_{max}) was determined by lysing the Quin 2-loaded cell with 0.1% Triton X-100. Minimum fluorescence (F_{min}) was obtained after addition of EGTA (4 mM) and sufficient Tris buffer to raise the pH to 8.0. Approximate values for cytoplasmic free Ca^{2+} were calculated from fluorescence readings F_x by the formula: $[Ca^{2+}]_x = 115$ nM $(F_x - F_{min}) / (F_{max} - F_x)$.

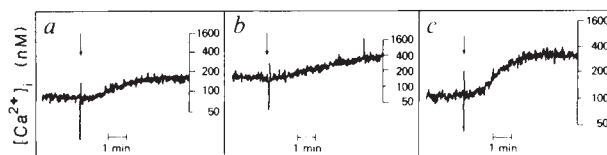


Table 1 IL-2 production by Thy-1.2 transfected Jurkat cell lines

CD3/Ti ⁺ , Thy-1 ⁺ and CD3/Ti ⁻ , Thy-1 ⁺ cell lines		Responder cell lines		
Stimulus		J14C9	18B3	1B3
Medium		800	3,200	500
TPA		3,100	2,800	3,500
OKT3 + TPA		157,300	3,600	2,300
Anti-Thy + TPA		69,100	7,300	1,700
Anti-Thy/Anti-Ig + TPA		118,500	4,400	6,000
A23187 + TPA		240,700	179,600	170,800

Ti α - and β -chain transfected cell lines		18B3	18B3 α 1.4	18B3 α 1.5	1B3	1B3 β 3E9	1B3 β B8
Stimulus							
Medium		700	300	700	6,600	7,600	7,600
TPA		500	5,000	2,900	8,300	13,600	9,100
OKT3 + TPA		900	75,500	80,400	4,900	66,300	77,400
Anti-Thy + TPA		600	30,400	61,400	5,400	69,100	45,700
A23187 + TPA		57,300	64,600	67,300	98,400	88,300	89,400

All cell lines were incubated in microtitre wells at 1×10^5 per well for 48 h with OKT3 ascites (1/1,000 dilution), a combination of rat anti-Thy-1 mAbs 3H11 and 3A7 (1/200 dilution of ascites), or A23187 (100 ng ml⁻¹), each in the presence of TPA (50 ng ml⁻¹). In the preincubation studies, the responder cell line was incubated at 4 °C for 1 h with a 1/100 dilution of a combination of 3H11 and 3A7 ascites, washed, and then cultured in the continuous presence of a mouse anti-rat κ -chain mAb (MAR 18.5, 1/100 dilution of ascites) and TPA (50 ng ml⁻¹). Supernatant fluids were assayed for IL-2 content by measurement of ³H-thymidine uptake by an IL-2 dependent cell line, CTLL. Data are expressed as mean ³H-thymidine incorporation of CTLL cells rounded to the nearest hundred. Neither phytohaemagglutinin (PHA), A23187, TPA, nor any of the mAbs used in this study had inhibitory or stimulatory effects on CTLL. All cell lines were maintained at 37 °C, 7.5% CO₂ in Dulbecco's modified Eagle's medium (DMEM) with 4.5 g l⁻¹ glucose, 10% NCTC109, 10% fetal calf serum (MA Bioproducts) non-essential amino acids (0.1 mM), sodium pyruvate (1 mM), penicillin (100 U ml⁻¹), streptomycin (100 μ g ml⁻¹), gentamycin (100 μ g ml⁻¹), glutamine (300 μ g ml⁻¹) and 2-mercaptoethanol (5×10^{-5} M).

Jurkat cells, but failed to induce IL-2 secretion by the CD3/Ti⁻ cell lines. It was therefore important to determine if this latter defect was a direct result of the lack of surface expression of CD3/Ti by these cells or if mutations other than in CD3/Ti components were present in these cell lines that prevented the induction of IL-2 production following stimulation through any cell-surface antigen.

Ohashi *et al.*¹³ demonstrated that the J.RT3-T3.5 cell line from which the 1B3 Thy-1.2⁺ transfectant was derived had a mutation in the functional Ti β -chain gene that can be corrected by transfection with a Ti β -chain complementary DNA clone, pT β Fneo, resulting in CD3/Ti expression on the cell membrane. We therefore transfected pT β Fneo into the CD3/Ti⁻, Thy-1.2⁺ cell line 1B3, which led to the expected reconstitution of CD3/Ti expression (Fig. 3b). The nature of the mutation responsible for the defect in CD3/Ti expression in the second Jurkat cell line (J.RT3-T3.1) was not established in the original studies describing this cell, as this line contained both 1.3- and 1.0-kilobase (kb) transcripts for the Ti β -chain and contained reduced levels

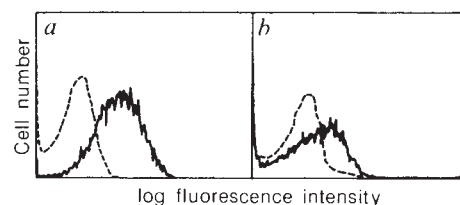


Fig. 3 OKT3 Staining of Ti α - and β -chain transfectants. The Ti α -chain transfectant of the 18B3 cell line, 18B3 α 1.4 (a) and the Ti β -chain transfectant of the 1B3 cell line, 1B3 β 3E9 (b) were stained with OKT3 ascites. FITC-conjugated second antibody in the absence of first antibody is shown as a control (broken line). **Methods.** The human T-cell receptor β -chain cDNA-containing vector pT β Fneo has been described previously. This cDNA encodes a protein identical to the Jurkat β -chain¹³. A full length murine Ti α -chain cDNA was isolated from the 2B4 murine T-cell hybridoma and subcloned into a plasmid containing the Friend spleen focus-forming virus long-terminal repeat and the *neo*^r gene to generate the p2B4 α Fneo vector¹⁶. pT β Fneo and p2B4 α Fneo were transfected to the CD3/Ti⁻, Thy-1⁺ Jurkat cells 1B3 and 18B3, respectively, as in the Fig. 1 legend, except that the selection medium consisted of G418 at 1 mg ml⁻¹.

of normal-sized α -chain transcripts. Because it had diminished levels of Ti α -chain gene transcripts, we reasoned that there might be a mutation in the functional α -chain gene. Transfection of the parental Jurkat line with a Ti α -chain cDNA derived from the murine T-cell hybridoma 2B4 leads to the appearance of surface molecules reactive with the anti-2B4 clonotypic antibody specific for the 2B4 α -chain¹⁶. This indicates that heterodimer formation occurs between human Ti β -chains and mouse Ti α -chains. If the defect in J.RT3-T3.5 and its Thy-1.2⁺ transfectant 18B3 resides only in the Ti α -chain gene, then transfection with the murine 2B4 α -chain cDNA should result in re-expression of CD3/Ti. This proved to be the case (T. Saito *et al.*, in preparation), as following transfection of the 18B3 cell line with the 2B4 α -chain cDNA construct, p2B4 α Fneo, several CD3/Ti positive clones were obtained. A representative profile of one such clone, 18B3 α 1.4, stained with OKT3 is shown in Fig. 3a. To determine whether re-expression of the CD3/Ti complex led to acquisition of the ability to produce IL-2 in response to anti-Thy-1 mAbs, two clones from each transfection were tested functionally. Both OKT3 and mAbs to Thy-1 could stimulate IL-2 production in the cells bearing the reconstituted CD3/Ti complex (Table 1, bottom), whereas the parental CD3/Ti⁻, Thy-1.2⁺ lines failed to produce IL-2 under the same conditions.

The results clearly demonstrate that the failure of CD3/Ti⁻, Thy-1⁺ Jurkat cells to produce IL-2 in response to anti-Thy-1 is due to a lack of surface expression of CD3/Ti, because transfection of the appropriate Ti α - or β -chain cDNA into the

defective cell lines restored CD3/Ti expression as well as responsiveness to stimulation by OKT3 and anti-Thy-1. It is not possible to predict from the present studies with Thy-1 whether T-cell activation through other cell-surface antigens (CD2, Tp44, TAP, Ly-6) also requires the expression of CD3/Ti. Indeed, the recent observation that a high concentration of mAb 9.3 can activate one CD3/Ti⁻ Jurkat mutant supports the view that activation through Tp44 does not involve CD3/Ti¹⁷.

One of the most striking findings in the present study is that the ability of anti-Thy-1 to increase $[Ca^{2+}]_i$ is independent of the expression of CD3/Ti. This finding is entirely consistent with our previous demonstration that anti-Thy-1 could induce a rise in $[Ca^{2+}]_i$, but failed to induce later events in B-cell activation (for example, immunoglobulin (Ig) secretion, unpublished observations), in CD3/Ti⁻, Thy-1.2⁺ transfected murine B lymphoma cells¹⁵. Our results are also consistent with the observations that mAbs against CD2 can induce rises in $[Ca^{2+}]_i$ in cells which are CD3/Ti⁻, CD2⁺^{18,19}. The functional significance of an increase in $[Ca^{2+}]_i$ without activation of genes leading to IL-2 production is entirely unknown. However, it is possible that the interaction of Thy-1 on the surface of a CD3/Ti⁻ cell with its natural ligand *in vivo* (perhaps in the thymus) leads to an increase in $[Ca^{2+}]_i$ and a subsequent differentiation event.

In contrast to most cell-surface proteins and glycoproteins that are anchored in the lipid bilayer of the cell membrane by hydrophobic peptide domains, Thy-1 has been shown to be attached to the cell surface via a phosphatidylinositol (PI) linkage²⁰. We have recently demonstrated that mAbs to Ly-6 activate murine T lymphocytes in a manner analogous to anti-

Thy mAbs¹¹ and that certain products of the *Ly-6* locus also appear to be linked to the cell membrane by a PI anchor (data not shown). It is thus tempting to speculate that Thy-1 may be a member of a unique group of cell-surface antigens that are anchored to the membrane by PI and that may have special roles in transmembrane signalling. The enhanced mobility afforded to Thy-1 (and Ly-6) by the PI linkage may facilitate the molecular interactions involved in signal transduction.

Received 27 October 1986; accepted 8 January 1987.

1. Chang, T. W., Kung, P. C., Gingras, S. P. & Goldstein, G. *Proc. natn. Acad. Sci. U.S.A.* **78**, 1805-1808 (1982).
2. Reinherz, E. L. *et al. Immun. Rev.* **81**, 95-129 (1984).
3. Haskins, K. *et al. J. exp. Med.* **157**, 1149-1169 (1983).
4. Meuer, S. C. *et al. Cell* **36**, 897-906 (1984).
5. Hara, T., Fu, S. M. & Hansen, J. A. *J. exp. Med.* **161**, 1513-1524 (1985).
6. Moretta, A., Pantako, G., Lopez-Botet, M. & Moretta, L. *J. exp. Med.* **162**, 823-838 (1985).
7. Gunter, K. C., Malek, T. R. & Shevach, E. M. *J. exp. Med.* **159**, 716-730 (1984).
8. MacDonald, H. R., Bron, C., Rousseaux, M., Horvath, C. & Cerottini, J.-C. *Eur. J. Immun.* **15**, 495-501 (1985).
9. Kroczeck, R. A., Gunter, K. C., Seligmann, B. & Shevach, E. M. *J. Immun.* **136**, 4379-4384 (1986).
10. Rock, K. L. *et al. J. exp. Med.* **163**, 315-333 (1986).
11. Malek, T. R., Ortega, G., Chan, C., Kroczeck, R. A. & Shevach, E. M. *J. exp. Med.* **164**, 709-722 (1986).
12. Weiss, A. & Stobo, J. D. *J. exp. Med.* **160**, 1284-1299 (1984).
13. Ohashi, P. S. *et al. Nature* **316**, 606-609 (1985).
14. Gunter, K. C., Kroczeck, R. A., Shevach, E. M. & Germain, R. N. *J. exp. Med.* **163**, 285-300 (1986).
15. Kroczeck, R. A., Gunter, K. C., Germain, R. N. & Shevach, E. M. *Nature* **322**, 181-183 (1986).
16. Saito, T., Weiss, A., Miller, J., Norcross, M. & Germain, R. N. *Nature* **325**, 125-130 (1987).
17. Weiss, A., Manger, B. & Imboden, I. *J. Immun.* **137**, 819-825 (1986).
18. O'Flynn, K. *et al. Eur. J. Immun.* **16**, 580-584 (1986).
19. June, C. H. *et al. J. clin. Invest.* **77**, 1224-1232 (1986).
20. Low, M. G. & Kincaid, P. W. *Nature* **318**, 62-64 (1985).

Modulation of *c-fos* gene transcription by negative and positive cellular factors

Paolo Sassone-Corsi & Inder M. Verma

Molecular Biology and Virology Laboratory, The Salk Institute, Box 85800, San Diego, California 92138, USA

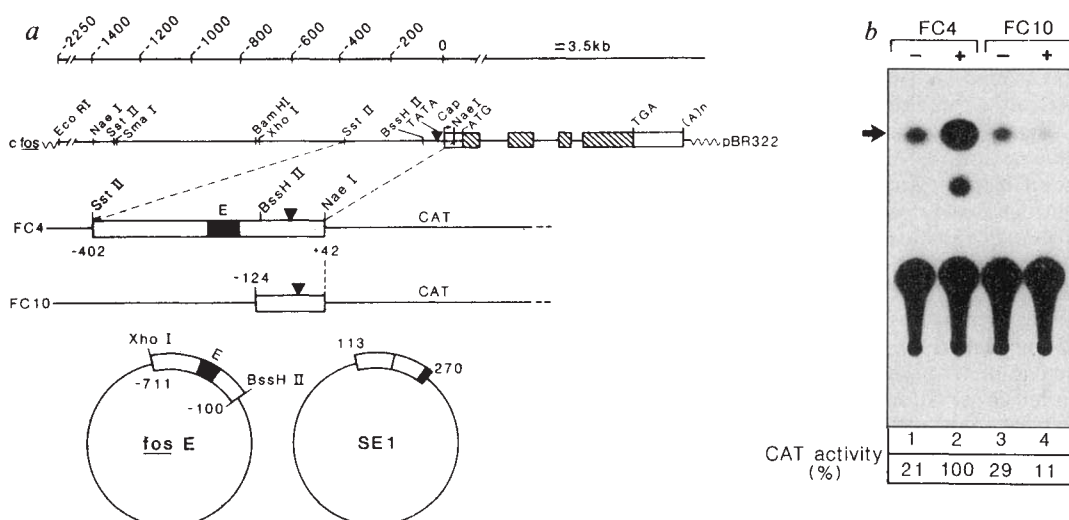
Regulation of transcription in eukaryotes is mediated by the specific interaction between cellular factors and promoter sequences¹. *Cis*-acting DNA sequences, frequently located upstream of the TATA box, have been implicated in modulating the expression of many genes². We are interested in the transcriptional regulation of proto-oncogenes because they may have a pivotal function in cell growth and differentiation^{3,4}. Expression of the proto-oncogene *c-fos* is induced in response to a variety of growth factors and differentiation-specific agents⁴. The viral cognate of the *c-fos* gene is the resident transforming gene of FBJ-murine osteosarcoma virus which causes bone tumours *in vivo* and cellular transformation *in vitro*^{5,6}. We report here that transcription of the human *c-fos* gene is modulated by negatively and positively acting cellular factors.

The *c-fos* gene promoter region has recently been delineated⁷⁻¹¹. In particular, an element essential for transcription activation has been localized between nucleotides -332 and -276 relative to the messenger RNA cap site⁷. It contains a region of hyphenated dyad symmetry and has properties similar to those described for a transcriptional enhancer^{7,8,11,12}. To understand further the regulation of *c-fos* expression, we designed *in vivo* experiments which could identify the presence of putative *trans*-acting cellular factors responsible for *c-fos* activation upon serum induction. To do so, we linked the *c-fos* promoter regulatory sequences to the coding sequences of the reporter gene chloramphenicol acetyl transferase (*CAT*)¹³. Two *fos*-*CAT* constructs were used in this study, FC4 and FC10 (see ref. 8 and Fig. 1a); FC4 contains the entire *fos* promoter region up to nucleotide -402; FC10 has a deletion to position

-124, so it lacks the transcriptional enhancer element. Figure 1b shows the induction (at least fivefold) observed when FC4-transfected NIH 3T3 cells are induced with serum (lanes 1 and 2), supporting the notion that the serum responsive element of the *c-fos* gene is located in the 5' flanking promoter region. The absolute levels of induction vary between 5- and 20-fold (compare also lanes 1 in Fig. 3a and b). In contrast, FC10, in which the serum responsive element has been deleted, is only weakly inducible. These findings are in general agreement with those previously reported^{7,14,15}. It has been shown by S₁-nuclease mapping that the initiation of transcription in these *fos*-*CAT* constructs is at the correct *c-fos* cap site^{7,10}.

Recent results indicate that *c-fos* gene transcription is very rapidly stimulated by serum even when the cells have been previously treated with protein synthesis inhibitors like cycloheximide or anisomycin^{16,17}. From this result and from the dramatic kinetics of the *c-fos* induction, it seemed possible that the *trans*-acting factor(s) required for *c-fos* gene transcription was present before serum treatment. To investigate this, we transfected increasing amounts of FC4 DNA into NIH 3T3 fibroblasts (Fig. 2). In cells which have been stimulated by serum after starvation, *c-fos* expression increases with increasing amounts of transfected FC4 DNA (lanes 1-4). However, *c-fos* expression obtained with 20 µg transfected DNA varies from 66% (as in Fig. 2) to over 80% (in other experiments) of that obtained with 30 µg of transfected DNA, indicating that the cellular factor(s) required for *c-fos* promoter activity is probably in limiting amounts in these cells. In contrast, in cells treated with only 0.5% serum (starved), the *c-fos* promoter activity was very low unless large amounts of FC4 DNA were used in the transfection (Fig. 2, lanes 5-8). This result lends strong support to the hypothesis that the positive cellular factor(s) required for *c-fos* transcription is present in fibroblasts before serum induction. Consequently, the reason for the decreased transcription of the *c-fos* promoter in serum-starved cells could be that a specific repressor molecule, which blocks the function of the positive factor(s) is present in these cells, or that the positive cellular factor(s) must undergo some modification during serum induction to activate the promoter.

Fig. 1 *a*, Structure of the human proto-oncogene *c-fos*. The transcription initiation site (Cap), the TATA box (▼), the exons (hatched boxes) and the position of several restriction sites in the upstream sequences are shown FC4 and FC10 are as described⁸. Briefly, FC4 is a *fos*-CAT fusion in which the *c-fos* promoter element between *Sst*II (−402) and *Nae*I (+42) has been cloned upstream from the CAT sequences¹². The filled segment in the *c-fos* upstream element locates the transcriptional enhancer previously described⁷. FC10



is a derivative of FC4, but contains a deletion that removes promoter sequences to position −124. Plasmids *fosE* and SE1 contain isolated promoter sequences from *c-fos* and SV40; *fosE* contains a *c-fos* promoter segment between nucleotides −711 (*Xho*I) and −100 (*Bss*HI) cloned in pUC19 (total size, 3.3 kilobases (kb)); SE1 (ref. 19) contains the SV40 enhancer from coordinates 113 to 270 inserted in pBR322 (total size, 2.7 kb). *b*, CAT activity of plasmids FC4 and FC10 after transfection into NIH 3T3 mouse fibroblasts. Plasmid DNA (5 µg) was transfected as described below; + and − indicate fibroblasts stimulated and starved by serum, respectively. The serum induction shown here using the FC4 plasmid is only an example; the stimulation varies from 5- to 20-fold.

Methods. Transfection of plasmid DNA into NIH 3T3 mouse fibroblasts was performed using the calcium phosphate co-precipitation technique¹⁹. Cells were passaged in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal calf serum (FCS). Cells were transfected at 70% confluence and exposed to the precipitate for 12–16 h. After washing, the medium was replaced with DMEM/0.5% FCS. After a further 48–72 h (at this stage we consider the cells starved by serum) cells to be stimulated were exposed to DMEM/15% FCS for a period of 30–40 min. CAT activity assays were performed as described¹². Quantification of CAT activity was performed by densitometer scanning of autoradiograms⁸, and is given as the ratio of the observed activity of the various samples to the maximal activity recorded in that particular experiment expressed as a percentage.

To distinguish between these two possibilities, we performed *in vivo* competition experiments¹⁸ in which increasing amounts of plasmid DNA containing promoter elements were co-transfected with a fixed amount of FC4. In cells stimulated by serum, neither pBR322 nor SE1 (a pBR322-based plasmid bearing the simian virus 40 (SV40) enhancer element¹⁹) sequences have any effect on *c-fos* transcription (Fig. 3*a*, lanes 2–4 and 8–10). On the other hand, sequences between nucleotide −711 and −100, decreased the transcriptional activity of the *c-fos* promoter (lanes 5–7). This result confirms the previous data (Fig. 2), indicating that the positive *trans*-acting factor(s) required for the *c-fos* promoter activity is in limiting amounts in serum-stimulated 3T3 fibroblasts. When the same competition experiments were performed in serum-starved cells, in which *c-fos* expression is very low (Fig. 3*b*, lane 1), an increase in CAT activity was detected with increasing amounts of the *fosE* plasmid competitor (lanes 5–7). This result supports the hypothesis that a repressor molecule blocks or diminishes the *c-fos* transcription in serum-starved fibroblasts. The repressor molecule is titrated out by increasing the concentration of the *c-fos* −711/−100 regulatory region but not by pBR322 sequences or by the SV40 enhancer element (Fig. 3*b*, lanes 2–4 and 8–10). Like the positive *trans*-acting factor, the repressor molecule is also in limiting amounts in the cell, because the titration starts with the addition of as little as 10 µg of *fosE* competitor. Direct comparison of the CAT activities between the experiments reported in Fig. 3*a* and *b*, indicates that after addition of 15 µg of *fosE* competitor, *c-fos* promoter activity in serum-starved fibroblasts reaches the same level of basal expression as observed in serum-stimulated cells (100% in lane 1, Fig. 3*a* and lane 7, Fig. 3*b*).

The results shown in Fig. 3*a* and *b* are measurements the levels of CAT activity but not the *c-fos* RNA levels. We subsequently performed RNase protection analysis of the transfected entire human *c-fos* gene. The basal transcriptional level after transfection of 2 µg of the *c-fos* plasmid in serum-stimulated or starved cells is shown in Fig. 3*c*, lanes 1 and 5. The induction

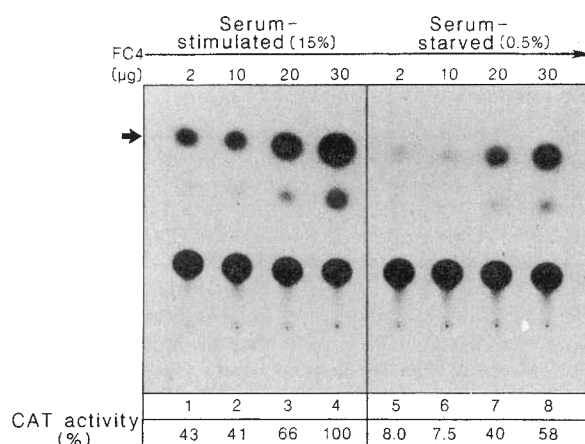


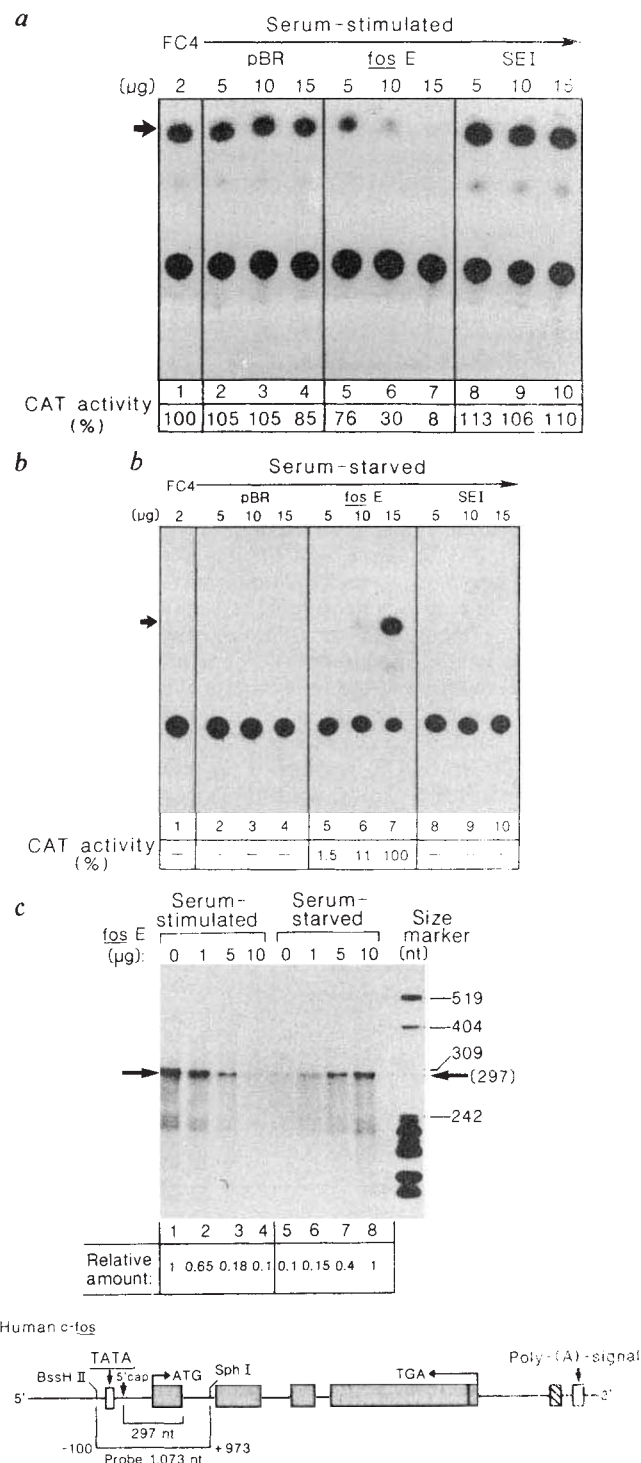
Fig. 2 CAT activity in 3T3 cells after transfection with increasing amounts of FC4 DNA. The total amount of DNA in each transfection was the same, being adjusted to 30 µg with pUC19 DNA when necessary. The transfection procedure was as in Fig. 1. The experiment shown here is representative of a series of transfections.

of *c-fos* transcription by serum is, as indicated, about 10-fold. The levels of *c-fos* RNA in serum-starved cells transfected with 10 µg of *fosE* are nearly identical (lane 8) to the basal level in serum-stimulated cells (lane 1). In serum-stimulated cells, addition of increasing concentrations of *fosE* plasmid decreases the levels of *c-fos* transcripts (lanes 2–4). Conversely, addition of *fosE* plasmid in serum-starved cells results in induction of the *c-fos* transcripts (lanes 6–8). These results are in close agreement with the data obtained with CAT assays (Fig. 3*a* and *b*).

When the same experiments as in Fig. 3*a* and *b* were performed using FC10 (Fig. 1), which lacks the *fos* enhancer element, instead of FC4, we were unable to detect any CAT activity in either serum-stimulated or starved cells (data not shown). Thus it appears that both positive and negative cellular factors

Fig. 3 Competition experiments in 3T3 fibroblasts stimulated by or starved of serum. *a*, Lane 1, CAT activity after transfection of 2 μ g of FC4 DNA in serum-stimulated 3T3 fibroblasts. Increasing amounts of pBR322 (lanes 2–4) and SE1 (lanes 8–10) competitor DNAs, as indicated, do not affect the FC4 expression. Increasing amounts of *fosE* (lanes 5–7) decreased the CAT activity of the co-transfected FC4. This experiment was repeated three times. *b*, Lane 1, CAT activity after transfection of 2 μ g of FC4 DNA in serum-starved 3T3 fibroblasts. Increasing amounts of pBR322 (lanes 2–4) and SE1 (lanes 8–10) competitor DNAs do not affect the FC4 expression. Increasing amounts of *fosE* (lanes 5–7) increased the CAT activity of the co-transfected FC4. This experiment was repeated three times. Quantification of the CAT activity, indicated below each lane, was by densitometer scanning of several autoradiograms. *c*, RNA protection analysis of the human *c-fos* gene transcripts: 2 μ g of a plasmid bearing the total human *c-fos* gene were transfected into serum-stimulated (lanes 1–4) and serum-starved (lanes 5–8) NIH 3T3 fibroblasts. Increasing amounts of *fosE* (lanes 2–4 and 6–8) were co-transfected, as indicated. Arrow, expected size of protected fragment.

Methods. Total cytoplasmic RNA was isolated and RNase mapping performed as described²⁹. Briefly, 30 μ g of RNA were hybridized with ³²P-labelled cRNA for 16 h at 45 °C in 80% formamide, 40 mM PIPES (pH 6.4), 400 mM NaCl and 1 mM EDTA. Following hybridization, samples were treated with a mixture of RNase A (40 μ g ml⁻¹) and RNase T1 (700U ml⁻¹) at 32 °C for 60 min. The samples were further incubated for 15 min at 37 °C after addition of 1% SDS and 50 μ g of proteinase K, extracted with phenol-chloroform, ethanol precipitated, and loaded on an 8% sequencing gel. The cRNA probe is 1,073 nucleotides (nt) long (from BssHII (-100) to SphI (+973)) and the expected size of the protected fragment is 297 nt. The intensity of the bands was measured by densitometer scanning of the autoradiogram.



interact with the upstream regulatory region. In addition, the lack of competition activity by the SV40 enhancer (SE1 in Fig. 3) suggests that the sequences required for the *c-fos* negative and positive regulation are not present in the SV40 element. As both SV40 and *c-fos* enhancer elements contain *core*-like sequences²⁰, the *c-fos* *core* element may not be directly responsible for the regulation observed in fibroblasts.

The results presented here indicate that the *c-fos* gene expression is regulated by both positive and negative trans-acting cellular factors. We show that the positive cellular factor(s) is present in fibroblasts, both when they are starved or stimulated with serum. The negative factor(s) is responsible for the *c-fos* transcriptional block in serum-starved fibroblasts, which is perhaps modified upon serum stimulation to allow efficient

expression. The modification(s) after serum treatment may involve variations in the affinity of the factor for specific sites on the promoter. We cannot exclude the possibility that positive and negative cellular components may be associated and that serum stimulation changes the properties of this complex. However the ability to titrate out the negative factor in serum-starved fibroblasts, using an excess of the upstream *c-fos* promoter element, is indicative of specific binding properties to promoter sequences. We propose that in serum-starved cells the negative factor(s) is associated with the *c-fos* promoter (or the promoter transcriptional complex), which results in the 'shut-off' of transcription. Upon serum stimulation, this factor(s) is modified and its affinity for the regulatory sequences decreases. (Alternatively, the positive factor is modified and its affinity for

the regulatory sequences increases.) In this model, the intracellular balance of the two classes of factors is crucial, as their affinity for *c-fos* regulatory sequences changes on serum treatment. This model will account for our competition results (Fig. 3), in which the negative element is titrated out, allowing the transcriptional complex to be activated. Our data, however, do not rule out the unlikely possibility that addition of high amounts of *fosE* in serum-stimulated cells can induce the negative-acting factor.

Other examples of negative regulation by *trans*-acting factors have been described, especially in viral systems. In undifferentiated embryonal carcinoma cells both Moloney murine sarcoma virus (Mo-MuSV) and polyoma enhancers are under negative regulation^{21,22}. The adenovirus-2 E1A products are the prototype of *trans*-repressors of transcriptional activity of enhancer sequences^{23,24}. We do not know if the repressor molecule described here has any similarity to the E1A-like proteins described in F9 cells²⁵. This is not a general feature of enhancers because SV40 and Rous sarcoma virus (RSV) promoter enhancer activity do not appear to be reduced in serum-starved fibroblasts (unpublished data). Examples of cellular regulatory regions that undergo a negative *trans*-regulation include the mouse immunoglobulin heavy-chain enhancer²⁶, the rat *insulin 1* promoter²⁷, and the interferon gene²⁸. The negative *trans*-regulation reported here is the first described for a proto-oncogene. A detailed understanding of the transcriptional regulation of cellular oncogenes is of great interest because of their possible role in differentiation and normal cell growth.

We thank Dr Vincent Raymond for his interest in this work and John Sisson for technical assistance. P.S.-C. is supported

by the American-Italian Foundation for Cancer Research and CNRS, and is on leave from the Laboratoire de Genetique Molculaire des Eucaryotes, CNRS, France. The work reported here was supported by the NIH and the American Cancer Society, Inc.

Received 18 August 1986; accepted 30 January 1987.

1. Sassone-Corsi, P. & Borrelli, E. *Trends Genet.* **2**, 215-219 (1986).
2. Serfling, E., Jasin, M. & Schaffner, W. *Trends Genet.* **1**, 224-230 (1985).
3. Muller, R. & Verma, I. M. *Curr. Topics Microbiol. Immun.* **112**, 73-115 (1984).
4. Verma, I. M. *Trends Genet.* **2**, 93-96 (1986).
5. Finkel, M. P., Biskis, B. O. & Jinkins, P. B. *Science* **151**, 698-701 (1966).
6. Miller, A. D., Curran, T. & Verma, I. M. *Cell* **36**, 51-60 (1984).
7. Treisman, R. *Cell* **42**, 889-902 (1985).
8. Deschamps, J. D., Meijlink, F. & Verma, I. M. *Science* **230**, 1174-1177 (1985).
9. Renz, M., Neuber, M., Kurz, C., Bravo, R. & Müller, R. *EMBO J.* **4**, 3711-3716 (1985).
10. Gilman, M., Wilson, R. N. & Weinberg, R. W. *Molec. cell. Biol.* **6**, 4305-4315 (1986).
11. Treisman, R. *Cell* **46**, 567-574 (1986).
12. Van Straaten, F., Muller, R., Curran, C., Van Beveren, C. & Verma, I. M. *Proc. natn. Acad. Sci. U.S.A.* **80**, 3183-3187 (1983).
13. Gorman, C. M., Moffat, L. F. & Howard, B. H. *Molec. cell. Biol.* **2**, 1044-1051 (1980).
14. Verma, I. M., Deschamps, J., Van Beveren, C. & Sassone-Corsi, P. *Cold Spring Harbor Symposium on Molecular Biology of Homosapiens* (1986) (in the press).
15. Deschamps, J. et al. *Cold Spring Harb. Symp. on Molecular Biology of Development* **50**, 733-745 (1985).
16. Greenberg, M. E., Hermanowski, A. L. & Ziff, E. B. *Molec. cell. Biol.* **6**, 1050-1057 (1986).
17. Mitchell, R. L., Henning-Chubb, C., Huberman, E. & Verma, I. M. *Cell* **45**, 497-504 (1986).
18. Scholer, H. & Gruss, P. *Cell* **36**, 403-411 (1983).
19. Sassone-Corsi, P., Wildemann, A. & Chambon, P. *Nature* **313**, 458-463 (1985).
20. Welher, H., König, M. & Gruss, P. *Science* **219**, 626-631 (1983).
21. Gorman, C. M., Rigby, P. W. J. & Lane, D. P. *Cell* **42**, 519-526 (1985).
22. Sassone-Corsi, P., Fromental, C. & Chambon, P. *Oncogene Res.* (submitted).
23. Borrelli, E., Hen, R. & Chambon, P. *Nature* **312**, 608-612 (1984).
24. Velcich, A. & Ziff, E. *Cell* **42**, 705-716 (1985).
25. Imperiale, M. J., Kao, H. T., Feldman, L. T., Nevins, J. R. & Strickland, S. *Molec. cell. Biol.* **4**, 867-874 (1984).
26. Borrelli, E. et al. *Proc. natn. Acad. Sci. U.S.A.* **83**, 2846-2849 (1986).
27. Nir, U., Walker, M. D. & Rutter, W. J. *Proc. natn. Acad. Sci. U.S.A.* **83**, 3180-3184 (1986).
28. Goodbourn, S., Burstein, H. & Maniatis, T. *Cell* **45**, 601-610 (1986).
29. Zinn, K., Dimaio, D. & Maniatis, T. *Cell* **34**, 865-879 (1983).

Transfection of the CD8 gene enhances T-cell recognition

Zlatko Dembić†, Werner Haast†, Rose Zamoyska*, Jane Parnes*, Michael Steinmetz† & Harald von Boehmer

Basel Institute for Immunology, Grenzacherstrasse 487, CH-4005 Basel, Switzerland

* Division of Immunology, Stanford University Medical Center, Stanford, California 94305, USA

Antibodies against CD8 or CD4 antigens can prevent T-cell functions induced by T-cell targets¹⁻³. As CD8 or CD4 antibodies can also initiate negative signals in T cells in the absence of appropriate targets⁴⁻⁶ it is not clear whether CD8 and CD4 molecules are directly involved in the interaction of T cells with their targets. In previous experiments we have introduced the T-cell receptor α - and β -chain genes from a CD8-positive cytolytic T cell specific for the antigen fluorescein (FL) and the H-2D^d molecule of the major histocompatibility complex (MHC) into a CD8-negative recipient cell. The CD8-positive donor cell lysed both FL-conjugated fibroblasts and lymphoblasts, which express relatively high and low amounts of H-2D^d molecules, respectively. In contrast the CD8-negative transfectant lysed FL-conjugated fibroblasts only⁷. Here we show that recognition of FL-conjugated lymphoblasts by the transfectant is enhanced by supertransfecting it with the CD8 gene.

The productively rearranged genes from the cytolytic T-cell clone BDFL1.1.3 (specificity: fluorescein (FL) and H-2D^d MHC antigen) have been transfected⁷ into the recipient cytolytic T-cell hybridoma SPH1.3 (specificity: *p*-sulphophenyldiazo-4-hydroxyphenyl acetic acid (SP) and H-2K^k MHC antigen). As shown in Fig. 1a,b and d, the BDFL1.1.3 donor, but not the

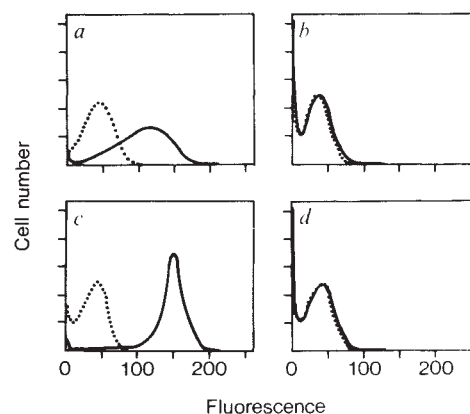


Fig. 1 Expression of CD8 antigen by the T-cell lines: a, BDFL 1.1.3; b, SPH 1.3; c, BD7-S17Ly2(2); d, BD7-S17. Cells were incubated with an excess of 53-6.7 antibodies ($100 \mu\text{g ml}^{-1}$) at 0°C for 30 min. After three washings in phosphate-buffered saline (PBS) the cells were incubated with fluorescein isothiocyanate (FITC)-coupled antibodies and washed twice before analysis with a fluorescence-activated cell sorter (solid lines). Broken lines represent cells incubated with FITC-coupled antibodies only.

SPH1.3 recipient or the BD7-S17 transfectant, expresses CD8 molecules recognized by the monoclonal antibody 53-6.7 (ref. 8). The BDFL1.1.3 donor but not the BD7-S17 transfectant lysed lipopolysaccharide (LPS)-induced DBA/2 lymphoblasts conjugated with FL (Fig. 3a; see also ref. 7). To force CD8 expression in the BD7-S17 transfectant we supertransfected BD7-S17 cells with the murine CD8 gene. To this end the CD8 gene was placed downstream of avian sarcoma virus long terminal repeat (ASV-LTR) enhancer elements in a vector also containing the hygromycin B resistance gene and, after linearization of the construct, introduced into BD7-S17 cells by electroporation (Fig. 2). Hygromycin B-resistant lines were screened for the expression of CD8 proteins by staining with the monoclonal CD8 antibody 53-6.7. About 10% of the hygromycin B-resistant clones expressed CD8 molecules (Fig. 1c). One of

† Present address: Central Research Units, Hoffman-La Roche & Co. Ltd, CH-4002 Basel, Switzerland.

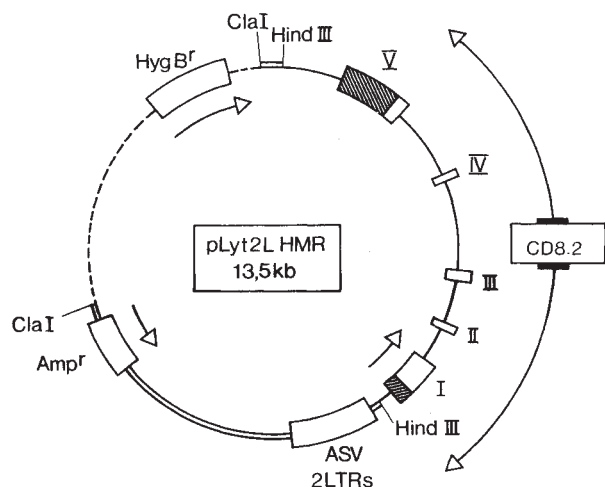


Fig. 2 CD8 gene construct for the transfection experiment: exons of the mouse CD8 (Ly-2) gene are shown as boxes with roman numerals; shaded boxes represent 5' and 3' untranslated regions; unshaded regions represent 5' and 3' untranslated regions; HygB^r, the hygromycin B resistance gene of pHMR272 (ref. 9); Amp^r, the β -lactamase gene of pBR322; ASV-2LTRs, two avian sarcoma virus long terminal repeats. Arrows, orientation of transcription. Double-lines, parts of plasmid pBR322/ASV-2LTR/tkA (ref. 10); broken line, segment derived from plasmid pHMR272; single line segment derived from the mouse CD8 gene⁹. Restriction enzyme sites used in the construction are shown.

Methods. A 5.4-kilobase (kb) *Hind*III fragment containing the mouse CD8 gene was subcloned into the *Hind*III site of pBR322/2LTR (derived from pBR322/ASV-2LTR/tkA by excising the HSV (herpes simplex virus)-*tk* gene with *Bam*HI, and converting the *Bam*HI site at the 3' end of the ASV-2LTRs into a *Hind*III site), giving pBR322/ASV-2LTR/Ly2. Then, pHMR272 (ref. 10) was linearized with *Cla*I (after the unique *Hind*III site was converted into a *Cla*I site) and inserted into the *Cla*I site of pBR322/ASV-2LTR/Ly2. The resulting plasmid (pLy2 LHM) which had lost the *Cla*I site 5' to the ampicillin-resistance gene was transferred into BD7-S17 cells by electroporation, using a commercially available machine (Dialog). The conditions were, $2.5-5 \times 10^7$ cells ml^{-1} in at 0°C, $25-50 \mu\text{g ml}^{-1}$ of *Cla*I-linearized DNA (added to the cells 10 min before pulsing), and 2 mg ml^{-1} of pronase (added to the mixture immediately before pulsing). Three pulses of 7 or 8 kV cm^{-1} using a 20 nF capacitor were delivered at 3-s intervals at room temperature. The cells were kept 10 min on ice, washed with medium and plated. Selection started three days after electroporation with 0.8 mg ml^{-1} of hygromycin B.

three selected CD8-positive clones exhibited cytolytic activity and lysed SP-coupled H-2K^k targets (Fig. 3b), whereas two others were no longer cytolytic. The cytolytic clone BD7-S17Ly2(2) was tested on a panel of LPS-induced lymphoblasts. In contrast to the CD8-negative BD7-S17 transfectant, the CD8-positive BD7-S17Ly2(2) clone lysed FL-conjugated DBA/2 lymphoblasts but lysed SP-conjugated or unconjugated DBA/2 lymphoblasts only at background levels. Thus, for three different types of DBA/2 targets, the specificity of the CD8 transfectant was very similar to that of the donor cell (Fig. 3a). FL- or SP-conjugated C57BL/6 (H-2^b) targets were not lysed by the CD8 transfectant (not shown). Unexpectedly, we found that on H-2^k lymphoblasts the specificity of the CD8-positive transfectant cell was different from both CD8-negative cell lines SPH1.3 and BD7-S17. The latter two lines lysed only SP conjugated H-2^k target cells but the former lysed all H-2^k targets irrespective of conjugation and much better than DBA/2 target cells (Fig. 3a, b). This new specificity revealed by transfecting the CD8 gene into BD7-S17 needs to be analysed in more detail.

CD8 antibodies interfered with the cytolytic activity of the BD7-S17 donor cell as well as the CD8-positive BD7-S17Ly2(2) cell when tested on fluorescein-conjugated DBA/2 LPS blasts. The cytolytic activity of the CD8-negative SPH1.3 cell on SP-conjugated CBA/2 LPS blasts was not affected by

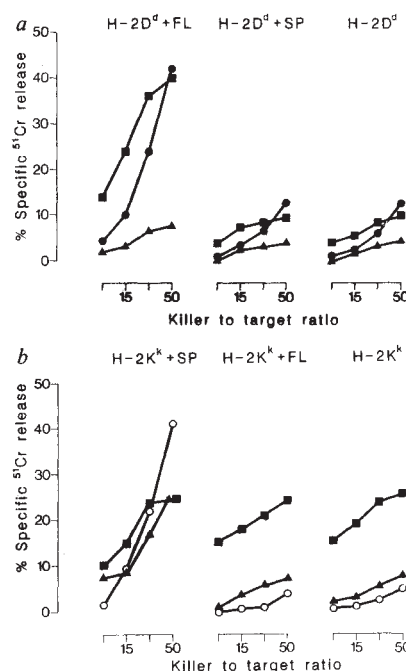


Fig. 3 Cytolytic activity of cytolytic T-cell clones on *a*, DBA/2 and *b*, CBA/J ⁵¹Cr-labelled LPS lymphoblasts. The cytolytic assay was conducted as described¹². Effectors were: *a*, ●, BD7-S17; ▲, BD7-S17; ■, BD7-S17-Ly2(2); *b*, ○, SPH1.3; ▲, BD7-S17; ■, BD7-S17-Ly2(2).

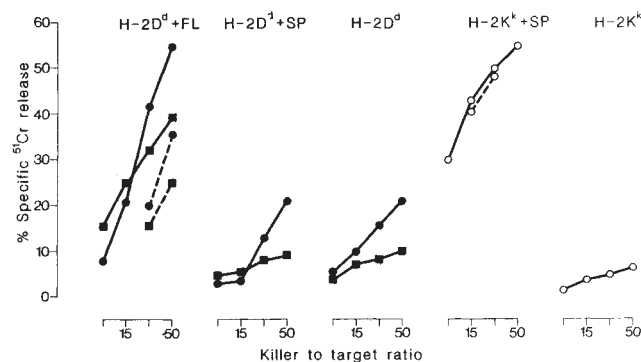


Fig. 4 Cytolytic activity of cytolytic T-cell clones on various target cells and inhibition by anti-CD8 antibodies. Assay conditions were as described¹². Effectors were: ●, BD7-S17; ■, BD7-S17-Ly2(2); ○, SPH1.3. Solid lines, no antibody; dotted lines, assay containing $100 \mu\text{g ml}^{-1}$ anti-CD8 antibody.

CD8 antibodies (Fig. 4). From these studies we conclude that the expression of CD8 molecules facilitates the interaction of cytolytic T cells with their targets. In addition the experiments further strengthen the conclusion⁷ that the specificity of T cells for antigen and polymorphic MHC determinants is solely mediated by the $\alpha\beta$ T-cell receptor.

We thank O. Nikkanen and K. Hafen for technical assistance, H.-P. Stahlberger for artwork and J. Hossmann for typing the manuscript.

Received 23 December 1986; accepted 10 February 1987.

- MacDonald, H. R., Glasebrook, A. L., Bron, C., Kelsoe, A. & Cerrotini, J. C. *Immun. Rev.* **68**, 89-115 (1982).
- Marrack, P. et al. *J. exp. Med.* **158**, 1077-1091 (1983).
- Greenstein, J. L., Malissen, B. & Burakoff, S. J. *J. exp. Med.* **162**, 369-374 (1985).
- Bank, I. & Chess, L. *J. exp. Med.* **162**, 1294-1303 (1985).
- Fleischer, B., Schrezenmeier, H. & Wagner, H. *J. Immun.* **136**, 1625-1628 (1986).
- Fleischer, B., Schendel, D. J. & von Steldern, D. *Eur. J. Immun.* **16**, 741-746 (1986).
- Dembić, Z. et al. *Nature* **320**, 232-238 (1986).
- Ledbetter, J. A. & Herzenberg, L. A. *Immun. Rev.* **47**, 63-90 (1979).
- Bernard, H. V., Krämmer, G. & Rövekamp, W. *Exp. Cell Res.* **158**, 237-243 (1985).
- Folger, K. R., Wong, E. A., Wahl, G. & Capecchi, M. R. *Molec. cell. Biol.* **2**, 1372-1387 (1982).
- Liew, C. W., Zamojska, R. & Parnes, J. R. *J. Immun.* **137**, 1037-1043 (1986).
- Haas, W., Pohlitz, H. & von Boehmer, H. *Eur. J. Immun.* **9**, 868-874 (1979).

A TATA box implicated in E1A transcriptional activation of a simple adenovirus 2 promoter

Lily Wu*, Debra S. E. Rosser†, Martin C. Schmidt* & Arnold Berk*‡

* Molecular Biology Institute and † Division of Cardiology, Department of Medicine, Center for the Health Sciences, University of California at Los Angeles, Los Angeles, California 90024, USA

Adenovirus E1A proteins stimulate transcription by RNA polymerases II and III from many promoters¹. The detailed mechanism of transcriptional activation (transactivation) by E1A proteins remains unclear, but genetic and biochemical results suggest that E1A products might act to stimulate the activity of cellular transcription factors¹. In this study, a detailed mutational analysis of the adenovirus *E1B* promoter was undertaken to define the DNA sequences required for proper basal transcription and E1A transactivation. Two key findings emerged: first the *E1B* promoter is an unusually simple RNA polymerase II promoter requiring only two sequence elements for proper regulation, the TATA box² and a binding site for transcription factor Sp1^{3,4}; and second only mutations in the TATA box interfere with E1A-transactivation, suggesting that E1A mediates its effect on this promoter through the TATA-box transcription factor.

Genetic studies of promoters activated by E1A have failed to locate a sequence element whose function is to mediate E1A transactivation⁵⁻⁸. Genetic and biochemical studies suggest that E1A, rather than being a transcription factor itself, up-regulates the activity of host-cell transcription factors (TFs)¹. To understand further the mechanism of E1A transactivation, we performed a systematic mutational analysis of the *E1B* promoter in which each mutant was analysed for its response to E1A transactivation. We anticipated that the *E1B* promoter might be an unusually simple promoter, as only 68 base pairs (bp) separate the *E1A* and *E1B* transcription units.

Initially, large deletions in the adenovirus genome were constructed in the region of the *E1B* promoter. Adenovirus mutants with deletions from nucleotide position -127 to -449 and from +6 to +68 relative to the start site of transcription produced the same steady-state concentration of E1B messenger RNA as a virus with a wild-type *E1B* promoter region when E1A protein was provided in *trans* (data not shown). In contrast, a mutant with a deletion from -127 to +16 did not produce any detectable specific *E1B* RNA. We concluded that all essential *E1B* promoter elements lie between -128 and +5.

To define these promoter elements in more detail, deletion and linker scan (LS)⁹ mutations were constructed by oligonucleotide-directed mutagenesis¹⁰ (see Figs 1a and 3 for names of these mutants). All the mutations were reintroduced into the viral genome¹¹. These *E1B* promoter mutant viruses additionally contain an *E1A* defect, allowing the analysis of basal *E1B* transcription in the absence of the stimulatory E1A protein. When needed, E1A proteins were provided in *trans* by coinfection with a helper virus, SI3B (supplied by J. Sambrook; N. Stow, unpublished data), containing a wild-type *E1A* region and a small deletion in *E1B*. The small deletion enables us to distinguish *E1B* RNA transcribed by SI3B from *E1B* RNA expressed from the *E1B* promoter mutants. All the deletions extended from -127, which is within the transcribed region of *E1A*, to varying distances towards the start site of *E1B* transcription (Fig. 1a). The steady-state quantity of *E1B* transcripts isolated from HeLa cells at 6 hours after infection was assayed by S₁ nuclease mapping^{12,13}. Correctly initiated *E1B* transcripts from the *E1B*-promoter mutant viruses protect a 219-base frag-

Table 1 Transcription of *E1B* and E1A transactivation of wild-type and mutant in *E1B* promoters

	<i>E1B</i> RNA (+E1A) (% wild type)	E1A induction (+/- E1A)
Deletion mutants		
Wild type (dl1500)	100	5.1
-74	78*	4.1*
-65	170	5.0
-55	87	5.6
-45	16	3.1
-35	13	4.6
-25 to +5	ND	—
Linker scan mutants		
Wild-type (dl1505)	100	5.6
LS-65/-56	54	3.2
LS-55/-46	24	6.1
LS-45/-36	20	4.4
LS-35/-26	4	1.4
LS-25/-16	20	5.3
LS-15/-6	80	6.4
LS-5/+5	56	4.2

Densitometry scanning results of unintensified autoradiograms. A Hoefer densitometer was used. Peak integration was performed by weighing. Each number represents averaged results from at least two independent experiments, except *, result of one experiment. ND, not detectable (<2%).

ment from S₁ digestion whereas the *E1B* transcripts from the SI3B virus (which provides wild type E1A proteins) produce a 145-base protected fragment (Fig. 1b). All these deletion mutants also produce an *E1A-E1B* fusion message due to the removal of the *E1A* polyadenylation signal¹⁴.

Deletion mutants -74, -65 and -55 did not produce less *E1B* RNA than the wild type (Fig. 1c, Table 1). In fact, -65 consistently produced 1.7-fold more RNA than the wild type. In contrast, a dramatic decrease in *E1B* RNA was observed from deletion mutants -45 and -35. The sequence GGGGCGGGGC (-48 to -39) is partially or completely deleted by these mutations. This sequence perfectly fits the consensus sequence for the binding site of transcription factor Sp1^{3,4}. Partial (-45) or complete (-35) deletion of this GC sequence results in a sixfold decrease in mRNA production. Deletion to -25 and further deletions also remove the TATA element. No detectable (<2% wild-type) *E1B* RNA was transcribed from these mutants at 6 hours after infection.

To determine if the transcription factor Sp1 indeed binds to the *E1B* promoter, a DNase footprinting experiment¹⁵ was performed (Fig. 2a). Sp1 protected an 18-nucleotide region of the transcribed strand, centred over the GGGGCGGGGC sequence. The *E1B* promoter also effectively competed with the SV40 (simian virus 40) early promoter for the binding of Sp1 (data not shown). These results indicate that Sp1 binds to the *E1B* promoter and disruption of this binding site greatly diminishes *E1B* transcription.

E1A transactivation of the deletion mutants -74 through -35 was examined in the experiment shown in Fig. 2b. Expression of E1A proteins resulted in a five- to sixfold increase in the *E1B* RNA level from the wild-type promoter as determined by comparing the 219-bp band in the - and + lanes. E1A also stimulated all the deletion mutants to a similar extent (Table 1). Thus, even mutants in which the Sp1 binding site has been deleted, -45 and -35, respond to E1A stimulation. These results suggest that E1A probably does not activate transcription through the Sp1 activity.

The *E1B* promoter was studied further using linker scan (LS) mutations (Fig. 3a). Amounts of *E1B* transcripts were compared in the wild type and LS mutants in the presence of E1A proteins (Fig. 3b; Table 1). The amount in LS-65/-56 was 54% of the

‡ To whom correspondence should be addressed.

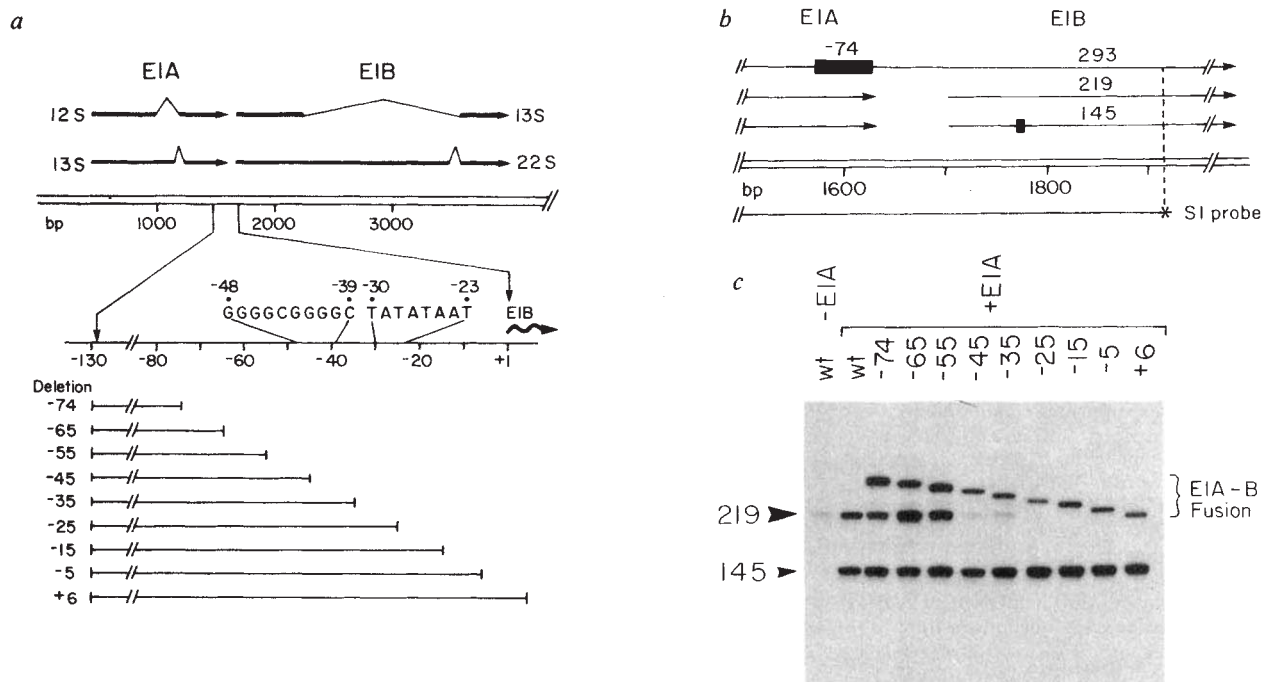


Fig. 1 *a*, Schematic diagram of the *E1B* promoter deletion mutants in relation to the *E1A* and *E1B* region of the adenovirus 2 (Ad2) genome. Lines at bottom, sequence removed in each mutant. All deletions extend from and include position -127 to varying 3' ends denoted by the mutant number. All the *E1B* promoter deletion mutations were introduced back into the viral genome with the dl1500 mutation³⁰ which prevents splicing of the 13S mRNA. The position of important regulatory elements in the *E1B* promoter, GGGGCGGGGC (-48 to -39) and TATA box (-30 to -23) are also shown. Above, the major *E1A* and *E1B* RNAs produced, with their respective sedimentation coefficients, aligned with the Ad2 viral sequence immediately below. *b*, DNA fragments expected to be protected in an S₁ nuclease mapping assay. The S₁ probe is a wild-type DNA fragment 5' end-labelled (*) at the *Bst*EII site at +219. Correctly initiated *E1B* RNAs from wild-type and *E1B* promoter deletion mutant viruses protect a 219-base fragment; the *E1B* RNA from the helper virus, S13B, protects a 145 base fragment. Every deletion mutant also produces an *E1A-E1B* fusion RNA. In the example shown, the *E1A-B* fusion RNA of deletion mutant -74 protects a 293-base fragment. *c*, Expression of *E1B* RNA from HeLa cells infected with wild-type and *E1B* promoter deletion mutant viruses at 6 h after infection. HeLa cells were infected at a multiplicity of infection (MOI) of 20 with wild-type viruses or ones carrying deletions in the *E1B* promoter. All *E1B* promoter deletion mutants also contained the dl1500 *E1A* mutation. '+E1A' indicates co-infection with the helper virus (MOI=20), S13B, to provide the E1A proteins *in trans*. Total cytoplasmic RNA was collected 6 h after infection and in each lane 100 µg of RNA was analysed by S₁ analysis with an excess of DNA probe. The autoradiogram of the S₁-protected fragments separated on a 5% denaturing acrylamide gel is shown.

Methods. All deletions were constructed by synthetic oligonucleotide-directed mutagenesis¹⁰ (except mutant -74 which was constructed by *Bal*31 digestion). The Ad2 *E1B* promoter region cloned in M13mp11 contains (with respect to *E1B* start at +1) -363 (*Xba*I) to +72 (*Sac*I) and an *Xho*I linker inserted at -127 (*Hpa*I), designated M13E1Bpr. Oligonucleotides (20-mers) having the first 10 bases complementary to the 3' end-point region of the deletion and the next 10 bases complementary to the *Xho*I linker (-127) were hybridized to the M13E1Bpr single-stranded template and, following the mutagenesis procedure of Zoller and Smith¹⁰, deletions with end points marked by an *Xho*I site were created. For example, deletion mutant -65 (M13E1BprD-65) has sequence from -127 to -66 inclusive removed. M13E1BprD 5' deletion clones were identified by preferential hybridization to the appropriate 20-mer oligonucleotide, and their sequences were confirmed by dideoxy sequencing. Similar methods generated 5' deletion clones -55, -45, -35, -25, -15 and +6. LS mutants were generated by recombining complementary 3' deletion (M13E1BprL) clones with the 5' deletion clones. The starting template for construction of M13E1BprL 3' deletions was the M13E1BprA clone which contained Ad2 *E1B* promoter sequence from -363 (*Xba*I) to +72 (*Sac*I); however, the *Sac*I site was converted to an *Xho*I site by blunt-end ligating an *Xho*I linker into the trimmed *Sac*I site. Using the M13E1BprA single-stranded template and specified mutagenic oligonucleotides as before, a series of M13E1BprL clones were generated. For example, M13E1BprL-65 contains *E1B* promoter sequence from -363 (*Xba*I) to -66 next to the *Xho*I site. By recombining the *Xba*I to *Xho*I fragment of M13E1BprL-65 into M13E1BprD-55, LS-65/-56 was generated. Using this strategy, the 10-bp *Xho* site replaced the respective sequence in LS-55/-46, LS-45/-36, LS-35/-26, LS-25/-16, LS-15/-6 and LS-5/+5. To generate mutant viruses, plasmids pBE1500 and pBE1505, which contain Ad2 sequence (1 to 3329) respectively from *E1A* mutant virus dl1500³⁰ and dl1505²⁹ positions, inserted into the *Pst*I site of pBR322 (modification of pBE5, N. Stow, personal communication) were used. Mutant dl1505 has an out-of-frame deletion following amino acid 22 in the *E1A* proteins. The *Xba*I-*Sac*I fragments from M13E1BprD deletion mutants were cloned into pBE1500. The *Xba*I-*Sac*I fragments from LS mutants were cloned into pBE1505. To generate mutant viruses, the linearized pBE1500/D and pBE1505/LS were cotransfected into 293 cells³¹ with dl309³² (an essentially wild-type variant with a single *Xba*I site at 1,336 in the Ad5 sequence) viral DNA cut at the unique *Xba*I site. Homologous recombination between cotransfected DNAs *in vivo* gave rise to mutant viruses (a modification of the method of Stow¹¹). Mutant viruses were identified by the diagnostic restriction digestion pattern of Hirt DNAs³³. Mutant viruses were plaque-purified at least twice and propagated in 293 cells. For viral RNA analysis, HeLa suspension cells were infected at a MOI of 20 and cytoplasmic RNAs were isolated at 6 h after infection as described³⁴. RNA (100 µg) was hybridized at 46 °C with excess DNA probe (1,336-1,918), singly 5' end-labelled at a *Bst*EII site (1,918). The S₁ digestion and gel electrophoresis conditions were as described²⁹.

wild-type amount; in LS-55/-46 and LS-45/-36, with substitution of the GGGGCGGGGC sequence, it was 20% of the wild-type level; and in LS-35/-26, in which the TATA sequence was mutated, it was only 4% of the wild-type level. In LS-25/-16, the amount was reduced to 20% of wild type. Sequences from -15 to -6 appear unimportant as this LS mutant produced essentially wild-type levels of *E1B* transcripts. Replacement of sequences over the cap site in LS-5/+5 decreased *E1B* transcription by 50%. It is interesting that LS-65/-56 and LS-55/-46 express a significant amount of *E1A-E1B* fusion message, because a GT-rich region is disrupted in these two LS mutants. G+U-rich sequences 3' of the AAUAAA polyadenylation signal are important for proper mRNA cleavage and polyadenylation of other primary transcripts¹⁶⁻¹⁸.

E1A transactivation was measured for each of the LS mutants (Table 1). With the exception of LS-35/-26, *E1A* increased *E1B* transcription from the wild-type and LS mutant promoters to a similar extent. In contrast, the *E1A* induction of LS-35/-26 was drastically reduced (Fig. 3b) to only 1.4-fold. This did not result from a lack of *E1A* function in the cells co-infected with LS-35/-26 and S13B as the intensity of the 145-base S₁-protected fragment from S13B was the same in all co-infections. Concentrations of *E1A* and *E2A* RNA were also measured in these same RNA preparations (data not shown) and were approximately equal in all the samples from the co-infected cells, confirming that the *E1A* transactivating function was expressed and functioning in all co-infected cells. The reduced *E1A*-transactivation of LS-35/-26 was quite different from all

Fig. 2 *a*, DNase footprinting of highly purified Sp1 on the *E1B* promoter. The DNA probe is the *E1B* promoter sequence, 5' end-labelled at +68 on the transcribed strand. Approximately 20 fmol of DNA probe was used per reaction³⁵. The Sp1 was affinity column purified³⁶. Lanes marked 0 contain no added protein. The next two lanes, marked Sp1, contain 0.1 and 0.2 µg of added protein respectively (left to right). Lanes marked G, G+A, C+T and C show Maxam and Gilbert sequencing ladders³⁷ using the same labelled DNA (transcribed strand). *b*, *E1B* transcription from wild-type viruses or ones with deletions in the *E1B* promoter in the presence or absence of the stimulatory *E1A* protein. The experimental conditions were as in Fig. 1c. Wt represents dl1500 13S *E1A*⁻ virus which is wild-type in the *E1B* promoter. In the analyses shown in (-) *E1A* lanes, RNAs were isolated from infections with *E1B* promoter mutants (MOI=20) which also contain the dl1500 *E1A* mutation; +*E1A* lanes indicate RNAs from the coinfections with the *E1B* promoter mutants and SI3B helper virus (MOI=20) to provide *E1A* functions in *trans*.

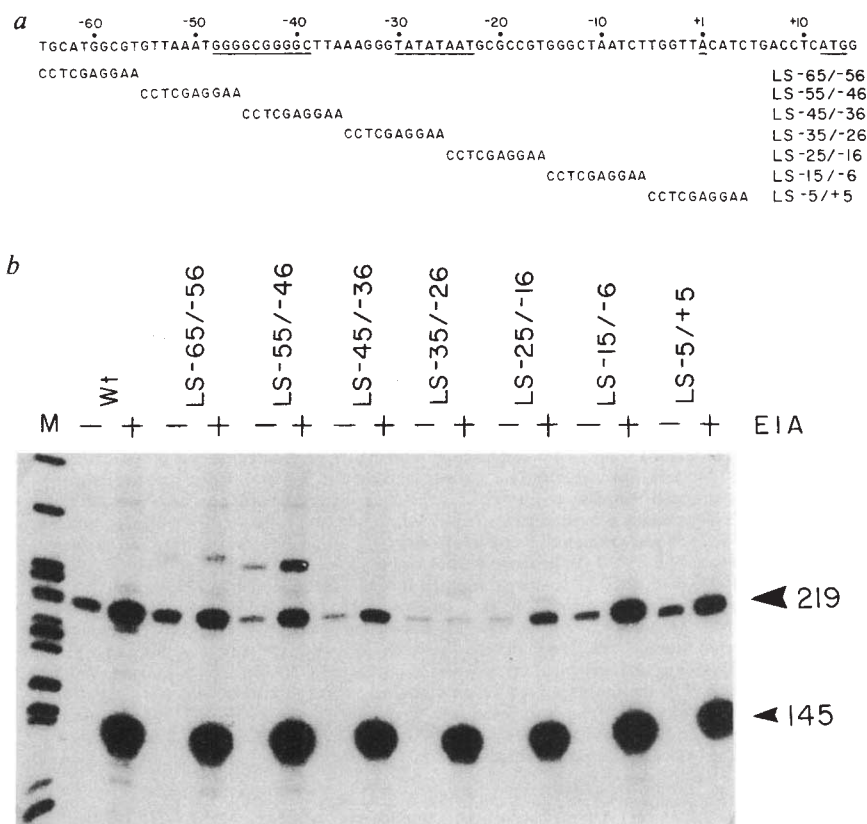
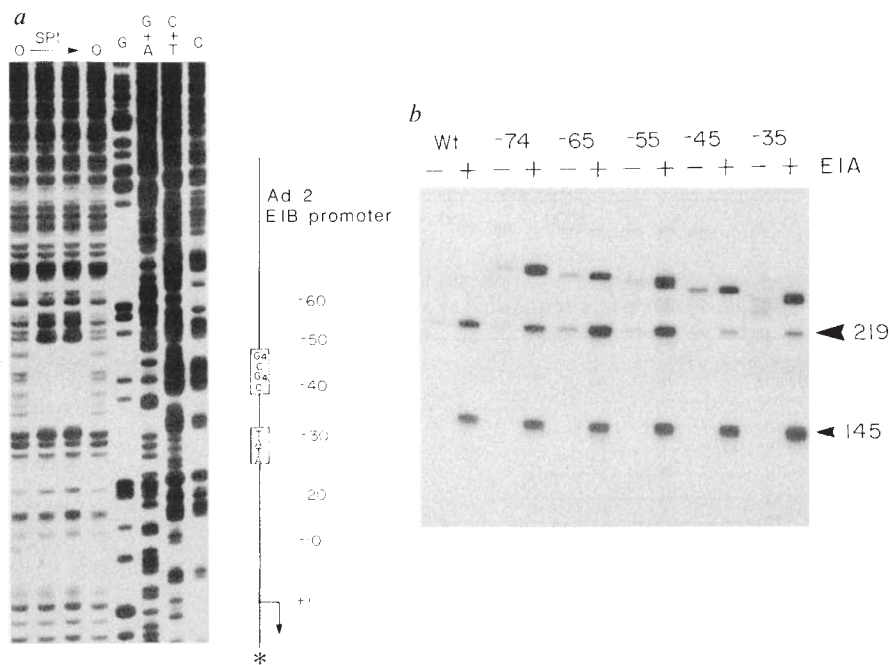


Fig. 3 *a*, The *E1B* promoter sequence substituted in each LS mutant. The top line shows the wild-type *E1B* promoter sequence with the transcription start site at +1. The important regulatory sequence elements are underlined. The wild-type sequence is substituted by the 10-bp *Xho* linker CCTCGAGGAA positioned directly below for each LS mutant. These *E1B* promoter LS mutations were introduced into the viral genome alone with a dl1505 *E1A* mutation²⁹ which truncates both large and small *E1A* proteins (see Fig. 1 methods). *b*, *S*₁ analysis of *E1B* RNA from wild-type *E1B* promoter and LS mutants. The experimental conditions were as in Fig. 1c. The wild-type *E1B* promoter virus is dl1505 (*E1A*⁻). *E1A* lanes represent absence (-) or presence (+) of the SI3B helper virus (providing *E1A* proteins in *trans*): 219-nucleotide products represent the correctly initiated *E1B* RNA transcribed from the wild-type or LS mutant viruses and 145-nucleotide products result from the helper SI3B *E1B* RNA. M, marker lane.

other *E1B* promoter deletion and LS mutants analysed. This result indicates that the TATA sequence substituted in LS-35/-26 is important in *E1A* transactivation of the *E1B* promoter.

Hearing and Shenk¹⁹ have shown that a sequence in the *E1A* upstream region, termed the *E1A* enhancer, stimulates *E1B* transcription about threefold. It is unlikely, however, that *E1A* proteins activate *E1B* transcription by affecting the activity of this *E1A* enhancer, as *E1A* protein provided in *trans* can still activate *E1A* transcription from a mutant virus with a complete deletion of the *E1A* enhancer¹⁹. In addition, an isolated Ad12 *E1B* promoter region can be transactivated by Ad12 *E1A* functions²⁰. However, all early viral promoters are stimulated three-

to fourfold by element II of the *E1A* enhancer²¹ and perhaps this is why the simple *E1B* promoter is a relatively active promoter in the context of the viral genome.

In most promoters examined, the TATA box is important for positioning the start site of transcription, and its deletion results in a depression of transcription by five- to tenfold^{2,22}. In the *E1B* promoter, the TATA element plays a crucial role in efficient transcription initiation *in vivo*. For the LS-35/-26 mutant, *E1B* RNA is reduced to only 4% of its wild-type level, but curiously the transcription start site is not altered (Fig. 3b). Most importantly, mutation of the TATA sequence greatly inhibited *E1A* transactivation of the *E1B* promoter. The TATA box is the

binding site for a host-cell RNA polymerase II transcription factor^{23,24}. Thus our results implicate the TATA-box transcription factor in E1A transactivation of the *E1B* promoter. E1A protein might increase the activity of the TATA-box factor itself, or increase the activity of a TF which must interact with the TATA-box factor to stimulate transcription.

In earlier work, Green *et al.*²⁵ found that TATA-box mutations significantly depress transcription of β -globin in transfected 293 cells (adenovirus-5 transformed human embryonic kidney cells which constitutively express E1A proteins). However, β -globin transcription was not detected following transfection into cells without E1A protein. Consequently, it was not possible to compare the effect of the TATA-box mutations on transcription in the presence and absence of E1A functions. Zajchowski *et al.*²⁶ reported that a mutation in a TATA-box-substitute sequence required for transcription initiation from a minor start site of the early *E2a* promoter depressed the level of E1A-transactivated transcription from the minor start site to a greater extent than the basal level. However, the same mutation had no effect on transcription from the major early *E2a* start site. Transfection-based assays of sequences required for E1A-transactivation of the *E3* promoter did not demonstrate a preferential effect of TATA-box mutations on transactivation²⁵. One possible explanation for this discrepancy is that the cell may contain more than one TATA-box TF and that E1A-functions stimulate the TF that interacts with *E1B* promoter, but not the TF that interacts with the *E3* promoter. Another possibility is that E1A functions may activate another TF that interacts with the complex *E3* promoter^{6,27} in addition to the TATA-box TF. As a result, *E3* TATA-box mutants would continue to respond to E1A-transactivation. That the *E3* promoter is more strongly transactivated than the *E1B* promoter²⁸ is consistent with this possibility. The activation of other TFs could explain E1A-transactivation of promoters which do not contain a TATA-box.

Greater understanding of E1A-transactivation will probably require further characterization of TFs in uninfected and infected cells. The *E1B* promoter may prove to be a very useful template in these studies because of its unusual simplicity.

We thank Dr Kathy Spindler for help in constructing viruses, Carol Eng, Kevin Beadles and Lisa Reizman for technical assistance, and Carol Crookshank for preparing this manuscript. This work was supported by the US PHS. L.W. was supported by the Medical Scientist Training Program Mandel Scholarship at the UCLA School of Medicine. M.S. was supported by a Postdoctoral Fellowship from the American Cancer Society. A.J.B. was supported by a Faculty Research Award from the American Cancer Society.

Note added in proof: We have recently constructed a mutant with a precise substitution of the TATA-box, LS-30/-23. It has a phenotype similar to that of mutant LS-35/-26 in that it is very poorly transactivated by E1A functions.

Received 10 October 1986; accepted 3 February 1987.

1. Berk, A. J. *A. Rev. Genet.* **20**, 45-70 (1986).
2. Breatnach, R. & Chambon, P. *A. Rev. Biochem.* **50**, 349-383 (1981).
3. Dynan, W. W. & Tjian, R. *Cell* **35**, 79-87 (1983).
4. Kadonaga, J. T., Jones, K. A. & Tjian, R. *Trends biochem. Sci.* **11**, 20-24 (1983).
5. Kingston, R. E., Kaufman, R. J. & Sharp, P. A. *Molec. cell. Biol.* **4**, 1970-1977 (1984).
6. Leff, T., Corden, J., Elkaim, R. & Sassone-Corsi, P. *Nucleic Acids Res.* **13**, 1209-1221 (1985).
7. Lewis, E. D. & Manley, J. L. *Molec. cell. Biol.* **5**, 2433-2442 (1985).
8. Murthy, S. C. S., Bhat, G. P. & Thimmappaya, B. *Proc. natn. Acad. Sci. U.S.A.* **82**, 2230-2234 (1985).
9. McKnight, S. L. & Kingsbury, R. *Science* **217**, 316-324 (1982).
10. Zoller, M. J. & Smith, M. *Meth. Enzym.* **100**, 468-501 (1983).
11. Stow, N. D. *J. Virol.* **37**, 171-180 (1981).
12. Berk, A. J. & Sharp, P. A. *Cell* **12**, 721-732 (1977).
13. Weaver, R. & Wiessman, C. *Nucleic Acids Res.* **7**, 1175-1193 (1979).
14. Montell, G., Fisher, E. F., Caruthers, M. H. & Berk, A. J. *Nature* **305**, 600-605 (1983).
15. Galas, D. L. & Schmitz, A. *Nucleic Acids Res.* **5**, 3157-3171 (1978).
16. Birnstiel, M. L., Busslinger, M. & Strub, K. *Cell* **41**, 349-359 (1985).
17. McLauchlin, J., Gaffney, D., Witton, J. L. & Clements, J. B. *Nucleic Acids Res.* **13**, 1347-1368 (1985).
18. Sadofsky, M. & Alwine, J. D. *Molec. cell. Biol.* **4**, 1460-1468 (1984).
19. Hearing, P. & Shenk, T. *Cell* **33**, 695-703 (1983).
20. Bos, J. L. & ten Wolde-Kraamwink, H. C. *EMBO J.* **2**, 73-76 (1983).
21. Hearing, P. & Shenk, T. *Cell* **45**, 229-236 (1986).
22. Grosschedl, R. & Birnstiel, M. L. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1432-1436 (1980).

23. Sawadogo, M. & Roeder, R. *Cell* **43**, 165-175 (1985).
24. Parker, C. S. & Topol, J. *Cell* **36**, 357-369 (1984).
25. Green, M. R., Triesman, R. & Maniatis, T. *Cell* **35**, 137-145 (1983).
26. Zajchowski, D. A., Boeuf, H. & Kedinger, C. *EMBO J.* **4**, 1293-1300 (1985).
27. Weeks, D. & Jones, N. *Nucleic Acids Res.* **13**, 5389-5402 (1985).
28. Nevins, J. R. *Cell* **26**, 213-220 (1981).
29. Osborne, T. F., Arvidson, D. N., Tyau, E. S., Dunsworth-Browne, M. & Berk, A. J. *Molec. cell. Biol.* **4**, 1293-1305 (1984).
30. Montell, C. & Berk, A. J. *Cell* **36**, 951-961 (1984).
31. Graham, F., Smiley, R. W. & Narin, R. J. *gen. Virol.* **36**, 59-72 (1977).
32. Jones, N. & Shenk, T. *Proc. natn. Acad. Sci. U.S.A.* **76**, 3665-3669 (1979).
33. Hirt, B. J. *molec. Biol.* **26**, 365-269 (1967).
34. Berk, A. J., Lee, F., Harrison, T., William, J. & Sharp, P. A. *Cell* **17**, 935-944 (1979).
35. Jones, K. A., Yamamoto, K. & Tjian, R. *Cell* **42**, 559-572 (1985).
36. Kadonaga, J. T. & Tjian, R. *Proc. natn. Acad. Sci. U.S.A.* **83**, 5889-5893 (1986).
37. Maxam, A. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 560-564 (1977).

Cyclin/PCNA is the auxiliary protein of DNA polymerase- δ

Rodrigo Bravo, Rainer Frank,
Patricia A. Blundell & Heather Macdonald-Bravo

European Molecular Biology Laboratory,
Postfach 10.2209, D-6900 Heidelberg, FRG

Identification of the cellular proteins whose expression is regulated during the cell cycle in normal cells is essential for understanding the mechanisms involved in the control of cell proliferation. A nuclear protein called cyclin of relative molecular mass 36,000 (M_r 36K), whose synthesis correlates with the proliferative state of the cell, has been identified in several cell types of human, mouse, hamster and avian origin¹⁻⁵. The rate of cyclin synthesis is very low in quiescent cells and increases several fold after serum stimulation shortly before DNA synthesis^{6,7}. Immunofluorescence and autoradiography studies have shown that the nuclear staining patterns of cyclin during S phase have a sequential order of appearance and a clear correlation can be found between DNA synthesis and cyclin positive nuclei⁷⁻¹⁰. The proliferating cell nuclear antigen (PCNA)^{8,11-13} and cyclin have many common properties and it has been shown that these two are identical^{12,14}. Recently a protein which is required by DNA polymerase- δ for its catalytic activity with templates having low primer/template ratios has been isolated from calf thymus¹⁵. We report here that cyclin and the auxiliary protein of DNA polymerase- δ are identical.

To continue studies on cyclin it was essential to raise antibodies against this protein. Cyclin was purified by two-dimensional gel electrophoresis from human MOLT-4 cells and injected as an antigen into rabbits. After several attempts, an antibody against cyclin was obtained which specifically precipitated cyclin from ³⁵S-methionine-labelled MOLT-4 cell extracts (Fig. 1). In separate analyses it was found that the antibody recognizes cyclin from various species, including some amphibians (data not shown).

During these studies an auxiliary protein for DNA polymerase- δ in its catalytic activity with templates having low primer/template ratios was described¹⁵. The reported M_r of 37K and pI of 4.5 of the protein were very similar to those of cyclin. As all previous observations suggested that cyclin was closely involved in DNA replication¹⁶ we examined cyclin and the auxiliary protein for a possible relationship.

As a first attempt to determine if cyclin and the auxiliary protein of DNA polymerase- δ were similar, we ran purified auxiliary protein and ³⁵S-methionine-labelled MOLT-4 cell extracts together in two-dimensional gel electrophoresis. The auxiliary protein was purified by a combination of conventional column chromatographic methods as described¹⁵. Cyclin and the auxiliary protein co-migrated on these gels (not shown).

To prove the identity between both proteins, we carried out immunoblotting analysis from two-dimensional gels of purified auxiliary protein and MOLT-4 cell extracts. Antibody against

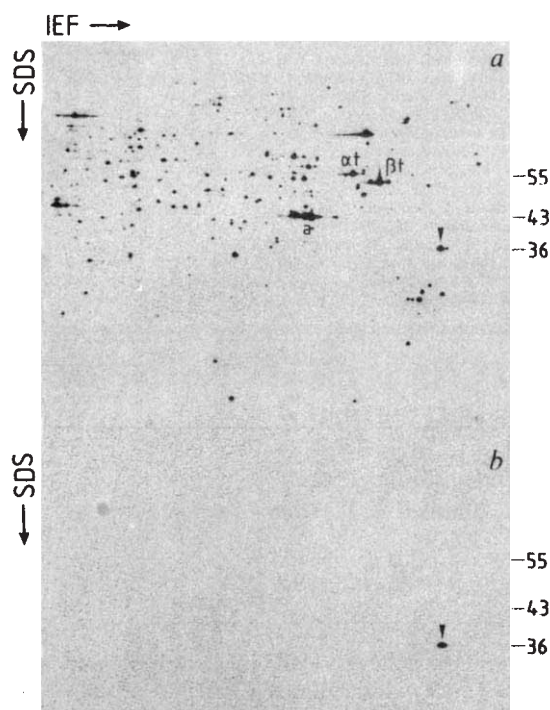


Fig. 1 Immunoprecipitation and two-dimensional gel analysis of cyclin. *a*, Exponentially growing MOLT-4 cells were labelled with ^{35}S -methionine for 16 hours. Total cellular proteins were extracted and analysed by two-dimensional gel electrophoresis. *b*, Total cellular extract from ^{35}S -methionine-labelled MOLT-4 cells was immunoprecipitated with anti-cyclin immunoglobulin G (IgG) from rabbit sera and analysed by two-dimensional gels. In *a*, αt is α -tubulin, βt is β -tubulin and *a* is actin; in *a* and *b*, cyclin is indicated with an arrowhead. Size markers, M_r in thousands.

Methods. In *a*, two-dimensional gel electrophoresis was as described⁵. Briefly, the first-dimension separations run to equilibrium (isoelectrofocusing, IEF) were performed on polyacrylamide gels (230×1.2 mm; 4% (w/v) containing 2% ampholytes (1.6%, pH 5 to 7; and 0.4%, pH 3.5 to 10) at 1,200 V for 20 hours. The second-dimension separations were done in a 15% acrylamide gel (25×25 cm) and were run at room temperature overnight. For antibody production the cell content of 31 of a MOLT-4 cell suspension (10^6 cells ml^{-1}) was used to prepare cyclin for immunization. After centrifugation cells were sonicated, treated with DNase and RNase²⁶ and passed several times through a narrow-gauge needle. Then the sample was lyophilized and resuspended in 7 ml of lysis buffer, and ^{35}S -methionine-labelled proteins from MOLT-4 cells were added to identify cyclin by autoradiography. One hundred and fifty two-dimensional gels were done as described⁵. Gels were immediately dried (without fixation) and exposed for 5 days. Cyclin was located, cut out from the gels and electroeluted. The antigen (~ 100 μg) was injected into rabbits using normal protocols. The rabbit sera were tested by immunoblotting. In *b*, immunoprecipitation of cyclin was as described²⁷ and immunocomplexes were isolated using *Staphylococcus aureus* as precipitant.

cyclin recognized the auxiliary protein in immunoblots (Fig. 2c). The specificity of the antibody had been previously established by the observation that the only protein detected in the immunoblots of MOLT-4 cell extracts was cyclin (Fig. 2b).

The amino-terminal sequence of cyclin had been reported¹⁷ and confirmed by sequencing a complete cDNA clone from MOLT-4 cells¹⁸. We determined the amino-terminal sequence of the auxiliary protein as final evidence of its identity with cyclin. For this analysis, the purity of the protein was confirmed by two-dimensional gel electrophoresis followed by silver staining (not shown).

The first ten amino-terminal residues of the auxiliary protein were identical to those of cyclin (Fig. 3), clearly establishing that the two proteins are the same. This is all the more interesting

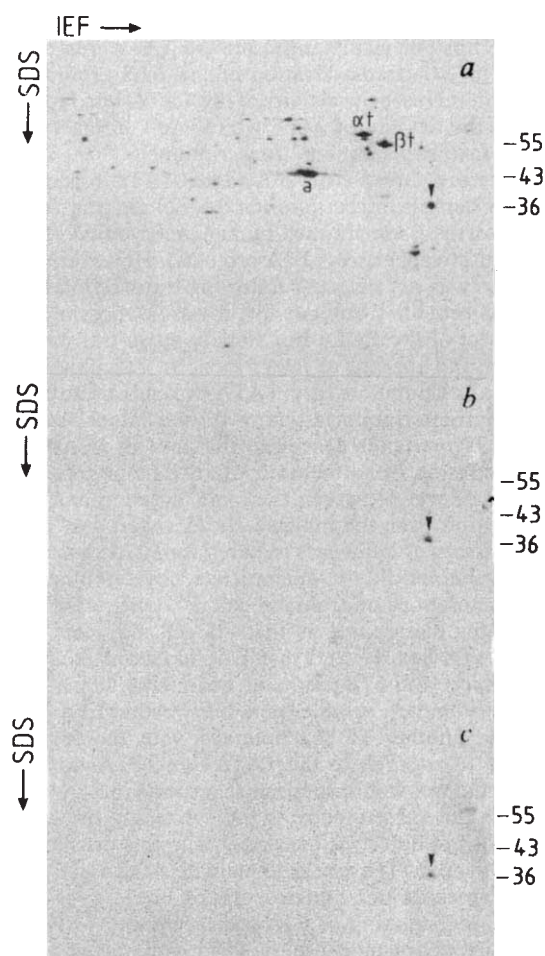


Fig. 2 Immunodetection of the auxiliary protein by blotting a two-dimensional gel with antibody against cyclin. *a*, Autoradiography of, and, *b*, immunodetection of cyclin by anti-cyclin rabbit antibody in, a total ^{35}S -methionine-labelled MOLT-4 cell extract. *c*, Immunodetection of auxiliary protein by anti-cyclin antibody. Size markers, M_r in thousands. Spots marked as in Fig. 1.

Methods. Proteins resolved by two-dimensional gels were transferred at 130 mA for 8 hours onto nitrocellulose filters as described²⁸. The filters were incubated in phosphate-buffered saline (PBS) containing 10% bovine serum and anti-cyclin antibody (diluted to 1/500) for 60 min at room temperature. The filters were then incubated with peroxidase-conjugated anti-rabbit IgG for 60 min in the same buffer. The immuno-complexes were visualized using diaminobenzidine as substrate.

in view of the finding that cyclin has a domain which is homologous to the α -helix/turn/ α -helix, the putative DNA-binding domain of several proteins¹⁸.

As previous observations had shown that cyclin synthesis was low in G_1 and high during S phase of the cell cycle¹⁹ we carried out an extended analysis of the rate of synthesis of cyclin during the cell cycle of HeLa cells. For this, cells were synchronized by mitotic shaking²⁰, plated and then labelled with ^{35}S -methionine for 1 hour every 2 hours for a period of 24 hours. DNA synthesis was followed in parallel by ^3H -thymidine incorporation and autoradiography (not shown). Figure 4 shows that the rate of synthesis of cyclin is very low in mid G_1 (3 hours after plating), increases during late G_1 (6 hours after plating), and reaches a maximum at 12–15 hours (mid-S phase). Thereafter the rate of synthesis decreases to basal levels during late S phase (19 hours after plating). These results confirmed and extended the previous findings on protein synthesis during the cell cycle of HeLa cells¹⁹.

It has been shown that induction of cyclin synthesis shortly precedes, and is independent of, DNA replication in quiescent

1 Met Phe Glu Ala Arg Leu Val Gln Gly Ser 10 Cyclin
 Met Phe Glu Ala Arg Leu Val Gln Gly Ser Auxiliary
 protein

Fig. 3 Amino-terminal sequences of the auxiliary protein and cyclin. The amino-terminal sequence analysis was performed in a gas-phase sequencer constructed and operated as described²⁹.

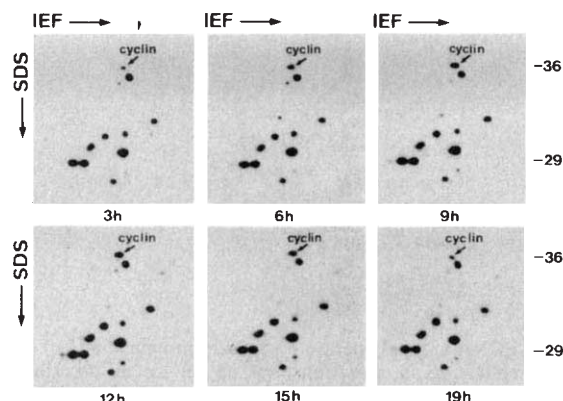


Fig. 4 Expression of cyclin during the cell cycle of HeLa cells. Synchronized cells were labelled with ³⁵S-methionine at different times after plating and radioactive polypeptides analysed by two-dimensional gel electrophoresis. In all cases 1×10^6 c.p.m. were applied to the gels. Only the area of interest is shown. Times indicate the beginning of the labelling period after plating the mitotic cells. Size markers, M_r in thousands.

Methods. Mitotic cells, obtained by mechanical detachment essentially as described²⁰, were plated in 2-cm² wells in Dulbecco's modified Eagle's medium (DMEM) supplemented with 5% fetal calf serum (FCS). Cells were labelled for one hour every two hours in DMEM without methionine in the presence of 1 mCi ml⁻¹ of ³⁵S-methionine supplemented with 5% dialysed FCS. After labelling, cells were extracted and polypeptides analysed as described⁵.

mouse fibroblasts stimulated to proliferate^{7,21}. This has been recently confirmed by the observation that the amount of cyclin messenger RNA following stimulation of quiescent fibroblasts increases in a similar fashion in the presence or absence of DNA synthesis inhibitors¹⁸.

An increase in the synthesis of DNA polymerase- α ^{22,23} but not of DNA polymerase- β ^{24,25} has been observed in proliferating cells and during the S phase. It is not known if the expression of DNA polymerase- δ changes during the cell cycle, but our studies suggest that it may. The knowledge that cyclin and the auxiliary protein of DNA polymerase- δ are identical, the availability of a specific antibody and the recent molecular cloning of this protein will be of great use for understanding its function in DNA replication.

We would thank Dr K. Downey for providing pure auxiliary protein and I. Benner for help with the manuscript.

Received 1 December 1986; accepted 4 February 1987.

- Bravo, R. & Graf, T. *Expl Cell Res.* **156**, 450-454 (1985).
- Bravo, R., Celis, A., Mosses, D. & Celis, J. E. *Cell Biol. int. Rep.* **5**, 479-489 (1981).
- Bravo, R. *et al. Expl Cell Res.* **136**, 311-319 (1981).
- Bravo, R., Fey, S. J. & Celis, J. E. *Carcinogenesis* **2**, 769-782 (1981).
- Bravo, R. *Proc. natn. Acad. Sci. U.S.A.* **81**, 4848-4851 (1984).
- Bravo, R. & Macdonald-Bravo, H. *EMBO J.* **3**, 3177-3181 (1984).
- Bravo, R. & Macdonald-Bravo, H. *EMBO J.* **4**, 655-661 (1985).
- Takasaki, J., Deng, J. S. & Tan, E. M. *J. exp. Med.* **154**, 1899-1909 (1981).
- Celis, A. & Celis, J. E. *Proc. natn. Acad. Sci. U.S.A.* **82**, 3262-3266 (1985).
- Madsen, P. & Celis, J. E. *FEBS Lett.* **193**, 5-11 (1985).
- Miyashita, K., Fritzler, M. J. & Tan, E. M. *J. Immun.* **121**, 2228-2234 (1978).
- Takasaki, J., Fischwild, D. & Tan, E. M. *J. exp. Med.* **159**, 981-992 (1984).
- Tan, E. M. *Adv. Immun.* **33**, 167-240 (1982).

- Mathews, M. B., Bernstein, R. M., Franza, R. R. & Garrels, J. I. *Nature* **309**, 374-376 (1984).
- Tan, C. K., Castillo, C., So, A. G. & Downey, K. M. *J. Biol. Chem.* **261**, 12310-12316 (1986).
- Bravo, R. *Expl Cell Res.* **163**, 287-293 (1986).
- Ogata, K., Ogata, Y., Nakamura, R. M. & Tan, E. M. *J. Immun.* **135**, 2623-2677 (1985).
- Almendral, J., Hübsch, D., Blundell, P., Macdonald-Bravo, H. & Bravo, R. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
- Bravo, R. & Celis, J. E. *J. Cell Biol.* **84**, 795-802 (1980).
- Terasima, T. & Tolmach, L. J. *Expl Cell Res.* **30**, 344-362 (1963).
- Macdonald-Bravo, H. & Bravo, R. *Expl Cell Res.* **156**, 455-461 (1985).
- Kozu, T., Seno, T. & Yagura, T. *Eur. J. Biochem.* **157**, 251-259 (1986).
- Thommes, R., Reiter, T. & Knippers, R. *Biochemistry* **25**, 1308-1314 (1986).
- Fry, M. in *Enzymes of Nucleic Acid Synthesis and Modification* (ed. Jacob, S. T.) 39-92 (CRC, Boca Raton, 1983).
- Chang, L. M. S. *Science* **191**, 1183-1185 (1976).
- Bravo, R. in *Two-dimensional Electrophoresis of Proteins* (eds Celis, J. E. & Bravo, R.) 4-36 (Academic, New York, 1984).
- Oppermann, H., Levinson, A. D., Varmus, H. E., Levintow, L. & Bishop, J. M. *Proc. natn. Acad. Sci. U.S.A.* **76**, 1804-1808 (1979).
- Burnette, W. N. *Analyt. Biochem.* **112**, 195-203 (1981).
- Gausepohl, H., Trosin, M. & Frank, R. in *Advanced Methods in Protein Microsequence Analysis* (eds Wittmann-Liebold, B. *et al.*) 149-160 (Springer, Heidelberg, 1986).

Functional identity of proliferating cell nuclear antigen and a DNA polymerase- δ auxiliary protein

Gregory Prelich*, Cheng-Keat Tan†, Matthew Kostura*, Michael B. Mathews*, Antero G. So†, Kathleen M. Downey† & Bruce Stillman*‡

* Cold Spring Harbor Laboratory, PO Box 100, Cold Spring Harbor, New York 11724, USA

† Departments of Medicine and Biochemistry and Center for Blood Diseases, University of Miami School of Medicine, Miami, Florida 33101, USA

The mechanism of replication of the simian virus 40 (SV40) genome closely resembles that of cellular chromosomes, thereby providing an excellent model system for examining the enzymatic requirements for DNA replication¹. Only one viral gene product, the large tumour antigen (large-T antigen), is required for viral replication², so the majority of replication enzymes must be cellular. Indeed, a number of enzymatic activities associated with replication and the S phase of the cell cycle are induced upon SV40 infection^{3,4}. Cell-free extracts derived from human cells, when supplemented with immunopurified SV40 large-T antigen support efficient replication of plasmids that contain the SV40 origin of DNA replication⁵⁻⁷. Using this system, a cellular protein of relative molecular mass 36,000 ($M_r = 36K$) that is required for the elongation stage of SV40 DNA replication *in vitro* has been purified and identified as a known cell-cycle regulated protein, alternatively called the proliferating cell nuclear antigen (PCNA) or cyclin⁸. It was noticed that, in its physical characteristics, PCNA closely resembles a protein that regulates the activity of calf thymus DNA polymerase- δ ⁹. Here we show that PCNA and the polymerase- δ auxiliary protein have similar electrophoretic behaviour and are both recognized by anti-PCNA human autoantibodies. More importantly, both proteins are functionally equivalent; they stimulate SV40 DNA replication *in vitro* and increase the processivity of calf thymus DNA polymerase- δ . These results implicate a novel animal cell DNA polymerase, DNA polymerase- δ , in the elongation stage of replicative DNA synthesis *in vitro*.

In an accompanying article⁸, we describe the fractionation of human 293 cell cytosolic extract (S100) into three components, each of which is required for complete replication of SV40 DNA, and the purification from one of these fractions of a protein that is required for efficient SV40 DNA synthesis *in vitro*. This replication factor was identified as PCNA, or cyclin, a protein shown previously to be cell-cycle regulated¹⁰⁻¹⁴. PCNA is preferentially detected by immunofluorescence in S-phase nuclei with human autoantibodies, raising the possibility that it is involved in cellular DNA replication or its control¹¹⁻¹³. The chromato-

‡ To whom correspondence should be addressed.

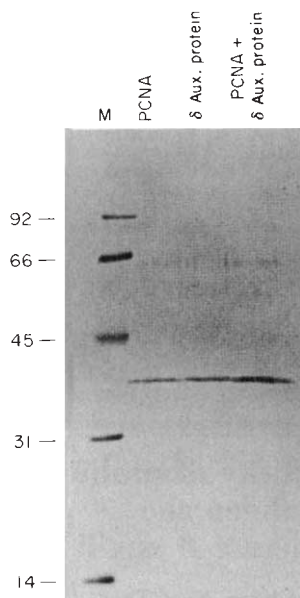


Fig. 1 Electrophoresis of PCNA and polymerase- δ auxiliary protein. PCNA was purified as described by Prelich *et al.*⁸ and the DNA polymerase- δ auxiliary protein was purified as described by Tan *et al.*⁹. The indicated proteins (0.5 μ g) were subjected to electrophoresis through a 12.5% SDS-polyacrylamide gel, which was stained with 0.1% Coomassie Brilliant Blue dye. Markers (M) at the left are M_r in K: phosphorylase B (92K), bovine serum albumin (66K), ovalbumin (45K), carbonic anhydrase (31K) and cytochrome *c* (14K).

graphic properties of PCNA, as well as its mobility during glycerol gradient sedimentation⁸, are remarkably similar to those recently described for a protein that enables fetal calf thymus DNA polymerase- δ to use template/primers containing long stretches of single-stranded template⁹. Indeed, during SDS PAGE (polyacrylamide gel electrophoresis), the two proteins co-migrate with an apparent M_r of 36K (Fig. 1). Therefore, we tested the ability of the polymerase- δ auxiliary protein to substitute for human PCNA in SV40 replication reactions. Both proteins supported replication to a similar extent (Fig. 2a). The authenticity of replication in this reconstituted system was evidenced by its dependence, both in the presence and the absence of PCNA, on large-T antigen and the SV40 origin, indicating that replicative, rather than repair synthesis was occurring (Fig. 2b).

In a reciprocal functional test for identity, we next examined the ability of human PCNA to stimulate the activity of purified calf thymus DNA polymerase- δ with poly(dA)/oligo(dT) (10:1) as template/primer. The results (Fig. 3a) demonstrate that both proteins stimulate polymerase- δ . To determine the effects of the auxiliary protein or PCNA on the processivity of the polymerase, the length of products synthesized in the presence and absence of the calf thymus auxiliary protein or human PCNA was determined by PAGE (Fig. 3b and c). Under conditions where <1 pmol of dTMP is incorporated per pmol of 3'-terminus, polymerase- δ produced only short products of <30 nucleotides. But when either the auxiliary protein or PCNA was included in the reaction the length of the polymerization products was dramatically increased, yielding products >200 nucleotides. We therefore conclude that PCNA is functionally identical to the polymerase- δ auxiliary protein and that this protein affects the processivity of polymerase- δ , although these data do not rule out a function in primer recognition as well. That the auxiliary protein increases polymerase- δ processivity strengthens the analogy between this protein and the β -subunit of the *Escherichia coli* polymerase III holoenzyme, as previously suggested⁹.

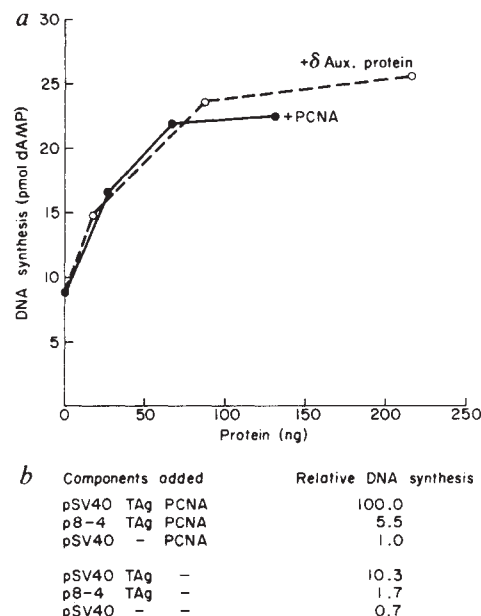


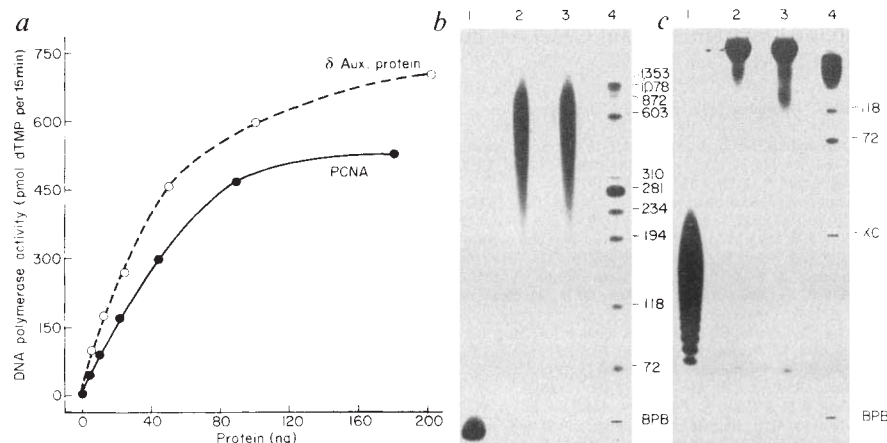
Fig. 2 PCNA and polymerase- δ auxiliary protein stimulate SV40 DNA replication. **a**, Reactions, set up on ice as described^{7,8}, contained fractions II, A and increasing amounts of either PCNA (solid line) or polymerase- δ auxiliary protein (broken line). After incubation of 1 h at 37 °C, the reactions were terminated by addition of EDTA to 10 mM, and the extent of ³²P-dAMP incorporation was determined as described⁷. **b**, Reactions were set up on ice containing cellular fractions II and A under standard conditions⁷ and included 300 ng of the indicated template DNA, 1 μ g of immunopurified large-T antigen, and 130 ng of PCNA as indicated. Plasmid p8-4 is an *ori*⁻ derivative of pSV40⁷. Reactions were incubated for 1 h at 37 °C, and then ³²P-dAMP incorporation was determined. TAg, large-T antigen.

Some patients with the autoimmune disease systemic lupus erythematosus (SLE) produce antibodies which recognize PCNA from a variety of mammalian species^{10,11}. These autoantibodies have been used to follow the distribution of PCNA (cyclin) throughout the cell cycle, and the preferential detection of PCNA in cells undergoing DNA synthesis led to speculation that PCNA functioned in cellular DNA replication^{12,13}. To determine whether the calf thymus polymerase- δ auxiliary protein is recognized by these human autoantibodies, we tested the ability of this purified protein to block the immunoprecipitation of ³⁵S-labelled HeLa-cell PCNA with antibodies from an SLE patient¹⁴ (Fig. 4). The calf thymus protein blocked the immunoprecipitation of the 36K PCNA band as efficiently as human PCNA, but minor co-precipitating bands were unaffected, indicating that the two proteins share a common epitope.

Experiments described here demonstrate that human PCNA and the calf thymus DNA polymerase- δ auxiliary protein are physically and antigenically related, and functionally indistinguishable in two assays. It was not obvious that the human PCNA and calf thymus polymerase- δ auxiliary protein would functionally substitute for each other because they have been isolated from different species. For example, the polymerase- α auxiliary proteins C1 and C2 exhibit species specificity, because C1 and C2 isolated from monkey and human cells are active only when assayed with polymerase- α isolated from the same species¹⁵.

The functional identity between PCNA and the δ -auxiliary protein raise the possibility that DNA polymerase- δ is involved in SV40 and host chromosomal DNA replication. Polymerase- δ is the most recently discovered and the least characterized of the mammalian DNA polymerases¹⁶⁻²⁰. Of the three other polymerases known, polymerase β has the characteristics of a repair enzyme, and the intracellular localization of polymerase γ impli-

Fig. 3 PCNA and polymerase- δ auxiliary protein increase the activity and processivity of DNA polymerase- δ . **a**, Incorporation of ^3H -dTMP into poly(dA)/oligo(dT) (20:1) with 1.4 units DNA polymerase- δ and increasing amounts of either PCNA or the δ -auxiliary protein. DNA polymerase- δ was purified to step 5 as described¹⁷. Assay conditions and preparation of δ -auxiliary protein were as in Tan *et al.*⁹. One unit of enzyme is defined as the incorporation of 1 nmol dNMP h⁻¹ with poly(dA-dT) as template/primer. **b**, **c**, Size of products synthesized by DNA polymerase- δ and either PCNA or δ -auxiliary protein with dA₅₀₀dT₁₀ (20:1) as template/primer. Each reaction contained 200 pmol of 3'-termini. Enzyme concentration was adjusted so that <1 pmol of dNMP was incorporated per pmol primer. **b**, PAGE (4% gel) of products in 7 M urea-TBE (80 mM Tris-HCl, 80 mM boric acid, 4 mM EDTA pH 8.3). Lane 1, products synthesized with 51 units of polymerase- δ alone (184 pmol [α - ^{32}P]dTMP incorporation in 10 min); lane 2, products synthesized with 0.12 unit of enzyme and 0.13 μg polymerase- δ auxiliary protein (60 pmol dTMP incorporated in 10 min); lane 3, products synthesized with 0.12 unit of enzyme and 0.13 μg PCNA (51 pmol dTMP incorporated in 10 min); lane 4, *Hae*III digest of ϕX174 , 5' [^{32}P] labelled. **c**, PAGE (20% gel) of products in 7 M urea-TBE. Lanes 1-4, as in **a**. BPB, bromophenol blue (comigrates with oligonucleotides of 8-10 residues on 20% gels); XC, xylene cyanol (comigrates with oligonucleotides of 28 residues on 20% gels). Gels were autoradiographed on Kodak XAR film with Dupont cronex intensifying screen.



cates it as the mitochondrial DNA polymerase (recently reviewed in ref. 21). DNA polymerase- α is widely assumed to be the sole polymerase involved in replicative DNA synthesis, partly because of the sensitivity of cellular and SV40 DNA synthesis *in vivo* to the antibiotic aphidicolin, a potent inhibitor of polymerase- α activity *in vitro*²¹. However, aphidicolin also inhibits polymerase- δ ¹⁷, making it necessary to re-evaluate some of the conclusions drawn from studies with this drug. Nonetheless, there is strong evidence implicating polymerase- α in replicative DNA synthesis. That polymerase- α is involved in SV40 DNA replication was demonstrated by the inability of human extracts immunodepleted of polymerase- α -DNA primase complex to support SV40 replication *in vitro*²². Replication activity was restored by the addition of human or monkey polymerase- α -primase, but not by the comparable enzymes from calf thymus or mouse cells, which are non-permissive for SV40 replication. That DNA polymerase- α also functions in cellular DNA replication is suggested by the ability of polymerase- α monoclonal antibodies to inhibit cellular DNA synthesis in isolated nuclei or when injected into cells^{23,24}. Furthermore, temperature-sensitive mouse FM3A cells, which contain a thermolabile DNA polymerase- α , are unable to replicate their DNA at the non-permissive temperature²⁵. Despite their similar response to aphidicolin, the α and δ polymerases appear to be distinct enzymes, differing in their chromatographic properties and template preferences^{17,20}. Polymerase- δ can also be distinguished from polymerase- α by its resistance to the inhibitor *p*-*n*-butylphenyl-dGTP(BuPdGTP)¹⁸⁻²⁰ and its association with an intrinsic 3'-to-5' exonuclease activity¹⁷, and polymerase α , but not polymerase δ , is tightly associated with a DNA-primase activity²¹. Moreover, the δ -auxiliary protein, now known to be the same as PCNA, stimulates polymerase- δ , but not polymerase- α ⁹.

Based on *in vitro* studies, we propose that SV40 DNA replication, and possibly cellular DNA replication, may require two distinct DNA polymerases. During SV40 replication, polymerase- α may recognize the SV40 origin of replication by direct or indirect interaction with the SV40 large-T antigen²⁶. The tightly associated DNA-primase activity could then synthesize primers, which would initially be elongated by polymerase- α . Because PCNA is required for the elongation stage of SV40 replication, we propose that polymerase- δ also functions during this stage of SV40 DNA replication. A function for DNA polymerase- δ in cellular DNA replication is more speculative, but the cell-cycle regulation of PCNA synthesis and localization

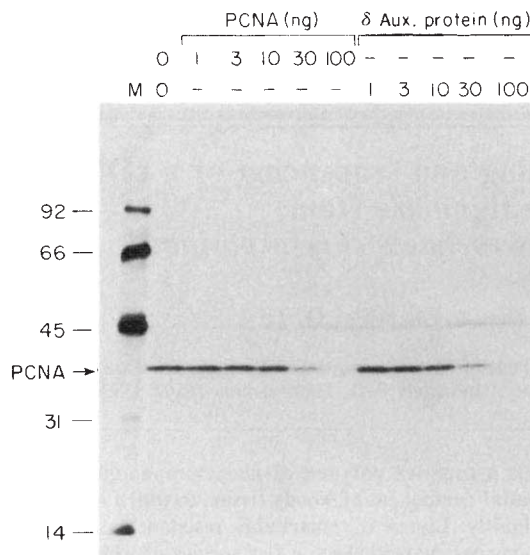


Fig. 4 Blocking immunoprecipitation of ^{35}S -labelled PCNA with purified PCNA or polymerase- δ auxiliary protein. The indicated amount of each protein was incubated for 15 min on ice with auto-antiserum obtained from an SLE patient¹⁴. Pre-cleared ^{35}S -labelled HeLa extracts were added and the reactions were further incubated for 15 min on ice. Immune complexes were precipitated with fixed *S. aureus* as precipitant and washed 6 times with 1 ml of (150 mM NaCl, 5 mM EDTA, 50 mM Tris pH 7.4, 0.5% Sodium deoxycholate, 0.5% Nonidet P40, 0.1% SDS). Precipitates were boiled and subjected to electrophoresis in a 12.5% SDS-polyacrylamide gel. The gel was treated with 1 M sodium salicylate for 30 min, then in H₂O for 20 min, dried and exposed to Kodak X-AR film. Markers are as in Fig. 1.

suggests a concomitant regulation of polymerase- δ function, just as occurs for polymerase- α activity²¹.

We thank Dr Robert M. Bernstein (Manchester Royal Infirmary) and Christopher C. Bunn (Royal Postgraduate Medical School, Hammersmith Hospital, London) for anti-PCNA serum, Richard Roberts for commenting on the manuscript and Barbara Faha for technical assistance. This work was supported by grants from the US Public Health Service (CA13106 and AM26206) and from the Rita Allen Foundation (to B.S.). G.P. is a predoctoral fellow and was partially supported

by training grant GM07964 and M.K. is a postdoctoral fellow supported by training grant CA09311, both from the US Public Health Service.

Received 8 December 1986; accepted 23 January 1987.

- DePamphilis, M. L. & Wassarman, P. M. in *Organization and Replication of Viral DNA* (ed. Kaplan, A. S.) 37-114 (CRC Press, Florida, 1982).
- Tegtmeier, P. J. *J. Virol.* **10**, 591-598 (1972).
- Sambrook, J. F. & Shatkin, A. S. *J. Virol.* **4**, 719-726 (1969).
- Kit, S., Piekarski, L. J. & Dubbs, D. R. *J. gen. Virol.* **1**, 163-173 (1967).
- Li, J. J. & Kelly, T. J. *Proc. natn. Acad. Sci. U.S.A.* **81**, 6973-6977 (1984).
- Li, J. J. & Kelly, T. J. *Molec. cell. Biol.* **5**, 1238-1246 (1984).
- Stillman, B. W. & Gluzman, Y. *Molec. cell. Biol.* **5**, 2051-2060 (1985).
- Prelich, G., Kostura, M., Marshak, D. R., Mathews, M. B. & Stillman, B. *Nature* **326**, 471-475 (1987).
- Tan, C., Castillo, C., So, A. G. & Downey, K. M. *J. biol. Chem.* **261**, 12310-12316 (1986).
- Miyachi, K., Fritzler, M. J. & Tan, E. M. *J. Immun.* **121**, 2228-2234 (1978).
- Tan, E. M. *Adv. Immun.* **33**, 167-240 (1982).
- Bravo, R. *Expl. Cell. Res.* **163**, 287-293 (1986).
- Celis, J. E. & Celis, A. *Proc. natn. Acad. Sci. U.S.A.* **82**, 3262-3266 (1985).
- Mathews, M. B., Bernstein, R. M., Franza, B. R. Jr & Garrels, J. I. *Nature* **309**, 374-376 (1984).
- Pritchard, C. G., Weaver, D. T., Baril, E. F. & DePamphilis, M. L. *J. biol. Chem.* **258**, 9810-9819 (1983).
- Byrnes, J. J., Downey, K. M., Black, V. L. & So, A. G. *Biochemistry* **15**, 2817-2823 (1976).
- Lee, M. Y. W. T., Tan, C. K., Downey, K. M. & So, A. G. *Biochemistry* **23**, 1906-1913 (1984).
- Lee, M. Y. W. T., Toomey, N. L. & Wright, G. E. *Nucleic Acids Res.* **13**, 8623-8630 (1985).
- Byrnes, J. J. *Biochem. biophys. Res. Commun.* **132**, 628-634 (1985).
- Crute, J. J., Wahl, A. F. & Bambara, R. A. *Biochemistry* **25**, 26-36 (1986).
- Fry, M. & Leob, L. A. *Animal Cell DNA Polymerases* (CRC press, Florida, 1986).
- Murakami, Y., Wobbe, C. R., Weissbach, L., Dean, F. B. & Hurwitz, J. *Proc. natn. Acad. Sci. U.S.A.* **83**, 2869-2873 (1986).
- Miller, M. R., Ulrich, R. G., Wang, T. S. & Korn, D. *J. biol. Chem.* **260**, 134-138 (1985).
- Kaczmarek, L., Miller, M. R., Hammond, R. A. & Mercer, W. E. *J. biol. Chem.* **261**, 10802-10807 (1986).
- Murakami, Y., Yasuda, H., Miyazawa, H., Hanaoka, F. & Yamada, M.-A. *Proc. natn. Acad. Sci. U.S.A.* **82**, 1761-1765 (1985).
- Smale, S. T. & Tijan, R. *Molec. Cell Biol.* **6**, 4077-4087 (1986).

Cloning and sequencing of a cDNA for a ligninase from *Phanerochaete chrysosporium*

Ming Tien & Chen-Pei D. Tu

Department of Molecular and Cell Biology, The Pennsylvania State University, University Park, Pennsylvania 16802, USA

Lignin is a complex polymer of phenylpropanoid subunits. It is an essential component of woody tissue, to which it imparts structural rigidity. Lignin is remarkably resistant to degradation by most microbes; nevertheless, a few species of white-rot fungi are able to catalyse its oxidation to CO₂. Its biodegradation is of great ecological significance because, next to cellulose, lignin is the most abundant renewable polymer on Earth. The first step in lignin degradation is depolymerization, catalysed by the lignin peroxidase isozymes (ligninases)¹⁻⁴. These isozymes are secreted, along with hydrogen peroxide (H₂O₂) by the fungus *Phanerochaete chrysosporium* Burds. under conditions of nutrient (nitrogen) limitation^{1,2}. Ligninases are not only important in lignin biodegradation, but are also potentially valuable in chemical waste disposal because of their ability to degrade environmental pollutants⁵. We have undertaken the cloning of the ligninase genes to understand further their regulation and enzymology. We report here the isolation and characterization of a ligninase complementary DNA clone with a full-length insert. The cDNA sequence shows that the sequence of the mature ligninase is preceded by a 28-residue leader, and the mature protein is predicted to have a relative molecular mass of 37,000 (*M_r*, 37K). Consistent with the classification of ligninase as a peroxidase certain residues thought to be essential for peroxidase activity can be identified and near these residues the ligninase shows homology with other known peroxidases. Our cDNA clone has also allowed us to show that expression of ligninase is regulated at the messenger RNA level.

When grown in liquid culture under conditions of nitrogen limitation, ligninase activity appears on days 4, 5 and 6 of

growth⁶. Consequently, the cDNA library was constructed⁷ with mRNA isolated from day-6 cultures. The primary library consisted of over 250,000 clones. Approximately 0.08% of the plaques were positive when screened with a polyclonal antibody which was raised against ligninase isozyme H8⁸. This percentage is in accord with the expected abundance from a protein which accounts for approximately 1% of the total cellular protein⁶. Phage DNA was isolated from 12 of the positive clones and the cDNA insert size was determined by agarose gel electrophoresis⁹. Six of these clones contained inserts larger than 1 kilobase (kb). Southern blot analysis showed that four of these hybridized to an oligonucleotide probe (5'-AAC GGC AAG ACT GTC GGC GAC GC-3' after residues Asn-Gly-Lys-Thr-Val-Gly-Asp-Ala) synthesized from a sequence of amino-acid residues near the N-terminus of the purified ligninase H8⁸. These four clones called λML-1, λML-2, λML-3 and λML-4 were then sequenced using the M13 dideoxy nucleotide chain termination method^{10,11}. Partial sequencing of the subcloned cDNA inserts in λML-2, λML-3 and λML-4 (350 base pairs (bp) from the 5' and 3' end) indicate that they are different copies of the same mRNA, but λML-3 is ~1 kb, whereas λML-2 and λML-4 are each ~1.4 kb. The sequence of λML-1 cDNA insert is similar to the others, but not identical.

The complete sequence of the λML-1 cDNA insert is shown in Fig. 1. The coding sequence is 1,122 bp of which 1,036 bp encode the mature ligninase. The experimentally determined N-terminal sequence of 23 amino acids, starting with Ala 1, was found in the deduced amino-acid sequence, thus unequivocally demonstrating λML-1 contains one of the ligninase genes. The sequence flanking the first ATG codon, 5'-GACATGG is consistent with the proposed eukaryotic translation initiation sequence of (A/G)NNATGG¹². The sequence in Fig. 1 also shows the presence of a 41-bp 5' non-coding sequence followed by an initiator codon and a 28-amino-acid leader sequence. Consistent with published results¹³, the leader sequence (-28 to -1) consists of hydrophobic amino-acid residues and ends with dibasic residues Lys-Arg. Dibasic residues have been found to be sites of endoproteolytic scission of precursor peptides such as the yeast α-factor¹⁴ and peptide hormones¹⁵. The presence of a leader sequence is also consistent with the ligninases being extracellular proteins¹. Translation of the sequence encoding the mature ligninase yields a protein of 345 amino acids. The calculated *M_r* of 37,072 is less than the experimental value of 42K³. The difference can be accounted for by glycosylation; the ligninase is a glycoprotein³. Only one potential *N*-glycosylation site can be identified (Asn 275) which follows the general rule of Asn-X-Thr/Ser¹⁶. There are, however, a total of 43 Ser/Thr residues which are potential sites for *O*-glycosylation. The deduced amino-acid composition also matches well with the determined amino-acid composition (data not shown). This further confirms the identity of λML-1 cDNA insert as a ligninase clone and supports the assignment of amino acids -28 to -1 as a signal peptide. The sequence AATAAA, implicated as a signal for polyadenylation¹⁷, is not present in the ligninase clone. However, a sequence of AAATAT, 12 bp upstream from the poly(A) tail may be the corresponding signal sequence for this fungal gene.

To study the regulation of the ligninase genes, mRNA was isolated from nitrogen-limited cultures (from day 1 to day 6 of growth). After electrophoresis, the RNA was subjected to Northern blot analysis using nick-translated λML-1 as a hybridization probe⁹ (Fig. 2a). The mRNAs encoding the ligninases are approximately 1.37 kb, which is very close in size to the cDNA insert (1,352 bp). The poly(A) RNAs were also assayed by *in vitro*¹⁸ translation and immunoprecipitation with a polyclonal antibody raised against ligninase H8. SDS-PAGE (polyacrylamide gel electrophoresis) of the immunoprecipitated proteins and fluorography revealed a set of multiple protein bands with *M_r* in the range 36-42K (Fig. 2b). The multiple bands are consistent with the *M_r* range of the ligninase isozymes (in the

Fig. 1 Nucleotide and predicted amino-acid sequence of the ligninase cDNA clone λ ML-1 from *P. chrysosporium*. The ligninase coding region is flanked by a 41-bp 5' non-coding and a 169-bp 3' non-coding region. The initiation codon is followed by a leader sequence (amino acids -28 to -1) consisting of predominantly hydrophobic amino acids. The underlined amino-acid sequences for residues 1-30 were obtained from amino-acid sequencing of the amino terminus, and matches with the deduced amino-acid sequence. Broken line above the sequence, the oligonucleotide probe (97-117) near the 5' end. Oligonucleotide primers (1,023-GAGCAGGCGTGTGCG-1,037; 171-CCACGGCGGCCAGTG-186; 332-GGGTGGAAACGCAGTC-318) were used at sites not corresponding to restriction sites for DNA sequencing. The DNA sequence of λ ML-1 cDNA was determined by the dideoxy method using [35 S]-deoxyadenosine-5'-(α -thio) triphosphate^{10,11}.

```

-41      *      TTTTTTCTTCAGTCCCACTCAGCACCAGCAACACAGCGGAC -1
1  ATG  GCC  TTC  AAG  CAG  CTC  TTC  GCA  GCT  ATC  TCT  CTC  GCT  CTC  TTG  CTC  TCG  51
-28 MET  Ala  Phe  Lys  Gln  Leu  Phe  Ala  Ala  Ile  Ser  Leu  Ala  Leu  Leu  Leu  Ser -12

52  GCT  GCG  AAC  GCG  GCT  GCG  GTG  ATC  GAG  AAG  CGC  GCC  ACC  TGT  TCC  AAC  GGC  102
-11 Ala  Ala  Asn  Ala  Ala  Ala  Val  Ile  Glu  Lys  Arg  Ala  Thr  Cys  Ser  Asn  Gly  6

103 AAC ACC GTC GGC GAT GCG TCG TCG TGC GCT TGG TTC GAC GTC CTG GAT GAT 153
7  Lys Thr Val Gly Asp Ala Ser Ser Cys Ala Trp Phe Asp Val Leu Asp Asp 23

154 ATC CAG CAG AAC CTG TTC CAC GGC GGC CAG TGC GGC GCT GAG GCG CAC GAG 204
24 Ile Gln Gln Asn Leu Phe His Gly Gly Gln Cys Gly Ala Glu Ala His Glu 40

205 TCG ATT CGT CTC GTC TTC CAC GAC TCC ATC GCA ATT TCG CCC GCC ATG GAG 255
41 Ser Ile Arg Leu Val Phe His Asp Ser Ile Ala Ile Ser Pro Ala Met Glu 57

256 GCA CAG GGC AAG TTC GGC GGC GGT GGT GCT GAC GGC TCC ATC ATG ATC TTC 306
58 Ala Gln Gly Lys Phe Gly Gly Gly Ala Asp Gly Ser Ile Met Ile Phe 74

307 GAC GAT ATC GAG ACT GCG TTC CAC CCT AAC ATC GGT CTC GAC GAG ATC GTC 357
75 Asp Asp Ile Glu Thr Ala Phe His Pro Asn Ile Gly Leu Asp Glu Val 91

358 AAG CTC CAG AAG CCA TTC GTT CAG AAG CAC GGT GTC ACC CCT GGT GAC TTC 408
92 Lys Leu Gln Lys Pro Phe Val Gln Lys His Gly Val Thr Pro Gly Asp Phe 108

409 ATC GCC TTC GCT GGT CGT GTC GCG CTC AGC AAC TGC CCT GGT GCC CCG CAG 459
109 Ile Ala Phe Ala Gly Arg Val Ala Leu Ser Asn Cys Pro Gly Ala Pro Gln 125

460 ATG AAC TTC TTC ACT GGT CGT GCA CCT GCT ACC CAG CCC GCT CCT GAT GGC 510
126 Met Asn Phe Phe Thr Gly Arg Ala Pro Ala Thr Gln Pro Ala Pro Asp Gly 142

511 CTT GTC CCC GAG CCC TTC CAC ACT GTC GAC CAA ATC ATC AAC CGT GTC AAC 561
143 Leu Val Pro Glu Pro Phe His Thr Val Asp Gln Ile Ile Asn Arg Val Asn 159

562 GAC GCA GGC GAG TTC GAT GAG CTC GAG CTT GTC TGG ATG CTC TCC GCG CAC 612
160 Asp Ala Gly Glu Phe Asp Glu Leu Glu Leu Val Trp Met Leu Ser Ala His 176

613 TCC GTC GCA GCG GTG AAC GAC GTC GAC CCG ACC GTC CAG GGT CTG CCC TTT 663
177 Ser Val Ala Ala Val Asn Asp Val Asp Pro Thr Val Gln Gly Leu Pro Phe 193

664 GAC TCG ACC CCC GGA ATC TTC GAC TCC CAG TTC TTC GTC GAG ACT CAG CTT 714
194 Asp Ser Thr Pro Gly Ile Phe Asp Ser Gln Phe Phe Val Glu Thr Gln Leu 210

715 CGT GGT ACC GCC TTC CCC GGC TCT GGT GGC AAC CAA GGC GAG GTC GAG TCG 765
211 Arg Gly Thr Ala Phe Pro Gly Ser Gly Gly Asn Gln Gly Glu Val Glu Ser 227

766 CCG CTC CCT GGC GAA ATT CGC ATC CAG TCC GAC CAC ACT ATC GCC GCG GAC 816
228 Pro Leu Pro Gly Glu Ile Arg Ile Gln Ser Asp His Thr Ile Ala Arg Asp 244

817 TCG CGC ACG GCG TGT GAA TGG CAG TCC TTC GTC AAC AAC CAG TCC AAG CTC 867
245 Ser Arg Thr Ala Cys Glu Trp Gln Ser Phe Val Asn Asn Gln Ser Lys Leu 261

868 GTC GAT GAC TTC CAG TTC ATC TTC CTC GCC CTC ACC CAG CTC GGC CAG GAC 918
262 Val Asp Asp Phe Gln Phe Ile Phe Leu Ala Leu Thr Gln Leu Gly Gln Asp 278

99  CCG AAC GCG ATG ACC GAC TGC TCG GAT GTT ATC CCG CAG TCC AAG CCC ATC 969
279 Pro Asn Ala Met Thr Asp Cys Ser Asp Val Ile Pro Gln Ser Lys Pro Ile 295

970 CCT GGC AAC CTC CCA TTC TCG TTC TTC CCC GCT GGC AAG ACC ATC AAG GAC 1020
296 Pro Gly Asn Leu Pro Phe Ser Phe Pro Ala Gly Lys Thr Ile Lys Asp 312

1021 GTT GAG CAG GCG TGT GCG GAG AGC CCC TTC CGA CTC TCA CCA CTC TCC CGG 1071
313 Val Glu Gln Ala Cys Ala Glu Ser Pro Phe Arg Leu Ser Pro Leu Ser Arg 329

1072 GCC CCG AGA CGT CCG TCC AGC GCA TCC CTC GCG CTC CCG GTG CTT AAA TGA 1122
330 Ala Pro Arg Arg Pro Ser Ser Ala Ser Leu Arg Leu Arg Val Leu Lys * 346

1123 TGCCATACAGAATACTCCTCAACCGACTGTAACGGTGCGCGGCTAACTCGTGACGGAATTCGGCT 1189

1190 TTACTAGATTTCATTATCATCTCTCTGCACCTAACTACGAATCTCATTCCTACTCTCTTACGAT 1256

1257 ATCTGCCGTGGGCTTATGAAATATCGGTGCACATCCAAAAAATAAAAAAAAAA 1311

```

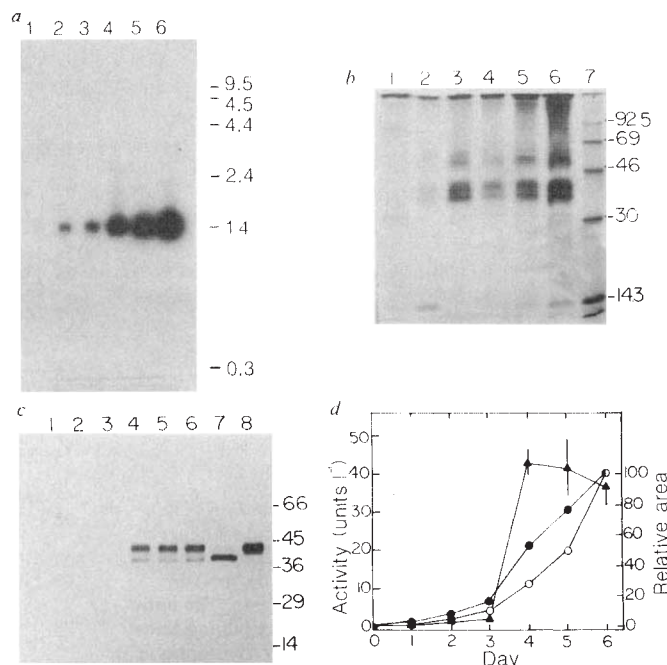
absence of glycosylation and processing of leader sequence)⁸. Another set of bands are also visible at $M_r \sim 55K$ (Fig. 2b). These minor bands correspond either to proteins with similar immunological epitopes or to contaminants in the antigen preparation used in antibody production. The appearance of mRNAs for the ligninases (Fig. 2a and b) parallels the appearance of the ligninase protein in the culture fluid as determined by SDS-PAGE and Western-blot visualization (Fig. 2c). Figure 2d shows the integrated area from the Northern blot and protein gel and extracellular ligninase activity plotted against days of growth. A temporal relationship is observed between the appearance of ligninase activity, mRNA and protein. Enzyme activity,

however, decreases after day 4 despite increases in both mRNAs and protein. This decrease in enzyme activity, without an associated decrease in total protein, has been previously observed, and is due to a decrease in the relative amounts of the more active ligninase isozymes⁶. These data clearly show that the ligninase gene expression is regulated at the level of mRNA production.

Some comparisons of the ligninase amino-acid sequence with those of other peroxidases are worthy of mention here. From kinetic¹⁹ and physical studies²⁰, ligninase can be categorized as a peroxidase. Peroxidases contain two histidine residues and an arginine which are essential for activity^{21,22}. One of the histidine

Fig. 2 Regulation of *P. chrysosporium* ligninase gene expression. *a*, Northern blot of poly(A) RNAs isolated from cultures of *P. chrysosporium* at various days of growth (lanes 1–6 correspond to days 1–6 respectively). Numbers to right of blot denote size of marker RNAs in kb. *b*, Fluorogram of an SDS-polyacrylamide gel of the immunoprecipitated *in vitro* translation products. Lanes 1–6 correspond to immunoprecipitated proteins from translation of poly(A) RNA isolated from days 1–6, respectively. Lane 7, size markers with numbers denoting M_r in thousands. *c*, Relative amounts of the ligninase enzymes at various days of growth. Concentrated extracellular fluid was subjected to SDS-PAGE and ligninase peptides were visualized by alkaline-phosphatase-conjugated immunoglobulin G (IgG) after Western transfer. Lanes 1–6 correspond to days 1–6 of growth. Lanes 7 and 8 contain the purified H2 and H8⁶, respectively. Size markers, M_r in thousands. *d*, Relative amount of the poly(A) RNA encoding the ligninase isozymes, obtained from the integrated areas of densitometer traces from the blot in *a* (●); the relative amount of the ligninase isozymes, obtained from the integrated areas of densitometer traces from the SDS-polyacrylamide gel (○); and the extracellular ligninase activity (▲) versus days of growth.

Methods. Poly(A) RNA was isolated from cultures of *P. chrysosporium* from days 1–6 of growth. Mycelium was isolated by passing cultures through cheesecloth followed by rapid freezing with liquid nitrogen and crushing in mortar and pestle. The cell powder was then suspended in guanidinium thiocyanate denaturing solution and centrifuged in CsCl²⁷. The RNA band was isolated and passed through an oligo(dT)-cellulose column to obtain the poly(A) RNAs²⁸. The poly(A) RNA was then used for either Northern blot hybridization (in *a*) or *in vitro* translation (in *b*). The Northern blot was from a 1.5% agarose gel containing formaldehyde⁹, hybridized with nick-translated λ ML-1 cDNA insert⁷; each lane contained 10 μ g of RNA. The RNA ladder was obtained from BRL. *In vitro* translation and immunoprecipitation was as described by Pelham and Jackson¹⁸ using 1 μ g of poly(A) RNA per incubation and antibody raised against the H8 isozyme. The H8 isozyme was purified as described by Tien and Kirk⁸, and the antibody isolated as described by Tu *et al.*²⁹. Immunoprecipitated proteins were visualized by SDS-PAGE³⁰ (10%) and fluorography²⁹. The ligninase isozymes were isolated from the extracellular fluid of day 1–6 cultures as described by Kirk *et al.*⁶. The culture fluid was then concentrated 20-fold by ultrafiltration. The Western blot³¹ in *c* was from gels containing 5 μ l of a 200-fold dilution of the concentrated extracellular fluid and 15 ng of H8 and H2 standards. Alkaline phosphatase-conjugated IgG was used as the second antibody. The ligninase activity of the extracellular fluid was measured by the rate of alcohol oxidation as described⁶. Results are mean and standard deviation of four cultures. A LKB 2202 Ultrosan laser densitometer was used to quantify band intensity.



The ligninase isozymes were isolated from the extracellular fluid of day 1–6 cultures as described by Kirk *et al.*⁶. The culture fluid was then concentrated 20-fold by ultrafiltration. The Western blot³¹ in *c* was from gels containing 5 μ l of a 200-fold dilution of the concentrated extracellular fluid and 15 ng of H8 and H2 standards. Alkaline phosphatase-conjugated IgG was used as the second antibody. The ligninase activity of the extracellular fluid was measured by the rate of alcohol oxidation as described⁶. Results are mean and standard deviation of four cultures. A LKB 2202 Ultrosan laser densitometer was used to quantify band intensity.

Distal histidine

Turnip 31 R M G A S T R L F F H D C F V N G C D⁵⁰.....153 A V G L S T R D M V A L S G A H T T Q S R¹⁷⁴
 CCP 41 G Y G P V L V R L A W H T S G T W D K H⁶¹.....159 Q R L N M D R E L V A L M G A H A L K T H¹⁸⁰
 HRP 31 R I A S T R L H F H D C F V N G C D⁵⁰.....155 V G L N R S S D L V A L S G A H T F K N Q¹⁷⁶
 Ligninase 36 A E I H E S T R L V F H D S I A T S P A⁵⁵.....161 A G E F D E L E L V W M L S A H S V I A A V N¹⁸²

Proximal histidine

Fig. 3 Comparison of *P. chrysosporium* lignin peroxidase (ligninase) and other peroxidases at regions near the proximal and distal histidine. Very little variation is observed for the location of the distal and proximal histidines of the various peroxidases. The distal histidines are located in the region of amino acids 42–52 whereas the proximal histidine is located in the region of amino acids 168–174 for turnip, cytochrome *c* (CCP) and horseradish (HRP) peroxidases. The ligninase predicted sequence contains eight histidine residues. The amino-acid residues of ligninase are aligned assuming a similar location of the His residues. Identical amino-acid residues are enclosed in solid boxes and residues of similar type are enclosed in dashed boxes. Amino-acid sequences for turnip, CCP and HRP were from refs 32, 13 and 33, respectively.

residues (proximal) is the axial ligand of the haem²³. The other histidine (distal) (along with the arginine) is proposed to be involved in charge stabilization during reaction of the haem with H₂O₂²³. Based on the X-ray crystal structure of cytochrome *c* peroxidase²¹ and kinetic studies²², the distal histidine is proposed to function as a general base catalyst in the H₂O₂ reaction (which results in oxidation of the haem by two electrons to form an intermediate referred to as compound I)²⁴. Thus the acid form of the distal histidine is thought to render the enzyme inactive and *pH* profiles for compound I formation show inflections around *pH* 5²⁵. The proximal histidine (His 175), the distal histidine (His 47), and the arginine (Arg 43) residues can all be identified in ligninase (Fig. 3). Comparing the amino-acid sequence of ligninase to turnip, cytochrome *c* and horseradish peroxidase, more homology is observed with the three latter peroxidases in residues flanking the proximal histidine. In contrast, the distal histidine with the Arg-Leu, four and three amino-acids upstream respectively, is conserved in all four peroxidases (Fig. 3). The sequence homology at the distal histidine would suggest that the mechanism of ligninase compound I

formation is similar to the other peroxidases. This, however, is unexpected because of the low *pH* optimum of ligninase catalysis³. (Our unpublished results indicate that the rate of the ligninase compound I formation does not show a *pH* dependency from *pH* 2 to *pH* 7.5.) These data raise some questions concerning the identity of the ionizable group affecting peroxidase compound I formation.

To identify which ligninase isozyme λ ML-1 corresponds to will have to await further protein and DNA sequencing. Evidence for more than one mRNA for the ligninases has been presented; heterogeneity is observed with the DNA sequences of two different clones. Furthermore, our Southern blots of digested genomic DNA also show multiple bands when hybridized with nick-translated λ ML-1 cDNA insert (not shown). These results are all consistent with the existence of multiple ligninase genes. Multiple ligninase genes have also been proposed by Zhang *et al.*²⁶. We are presently sequencing the other clone and plan to study the regulation of these ligninase genes. Finally, comparisons of amino-acid sequence near the active site of ligninase to other peroxidases have yielded information

inconsistent with proposed kinetic mechanisms. This raises some interesting questions on the chemical mechanism of compound I formation.

We thank Hui-Chen J. Lai and Nan-qian Li for their help throughout the study and R. C. Hardison and T. H. Kao for reviewing the manuscript. Amino-acid sequence and composition were determined at University of California at Los Angeles Medical School and University of Wisconsin Biophysics Laboratory, respectively. This work was supported in part by a Biomedical Research Support Grant from The Pennsylvania State University and a USDA grant to M.T. M.T. is recipient of a Presidential Young Investigator Award from NSF. C.-P.D.T. is recipient of a Research Career Development Award from USPHS. The sequence data in this publication have been submitted to the EMBL/GenBank Libraries under the accession number Y00262.

Received 25 November 1986; accepted 29 January 1987.

1. Tien, M. & Kirk, T. K. *Science* **221**, 661-663 (1983).
2. Glenn, J. K., Morgan, M. A., Mayfield, M. B., Kuwahara, M. & Gold, M. H. *Biochem. biophys. Res. Commun.* **114**, 1077-1083 (1983).
3. Tien, M. & Kirk, T. K. *Proc. natn. Acad. Sci. U.S.A.* **81**, 2280-2284 (1984).
4. Gold, M. H., Kuwahara, M., Chiu, A. A. & Glenn, J. K. *Archs Biochem. Biophys.* **234**, 353-362 (1984).
5. Bumpus, J. A., Tien, M., Wright, D. & Aust, S. D. *Science* **228**, 1434-1436 (1985).
6. Kirk, T. K., Croan, S. C., Tien, M., Murtagh, K. & Farrel, R. *Enzyme Microb. Technol.* **8**, 27-32 (1985).
7. Gubler, U. & Hoffman, B. J. *Gene* **25**, 263-269 (1983).
8. Tien, M. & Kirk, T. K. *Meth. Enzym.* (in the press).
9. Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning, a Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).
10. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
11. Messing, J. *Meth. Enzym.* **101**, 20-77 (1983).
12. Kozak, M. *Nucleic Acids Res.* **9**, 5233-5252 (1981).
13. Kaput, J., Goltz, S. & Blobel, G. *J. biol. Chem.* **257**, 15054-15058 (1982).
14. Julius, D., Schekman, R. & Thorner, J. *Cell* **36**, 309-318 (1984).
15. Bell, G. I., Santerre, R. F. & Mullenbach, G. T. *Nature* **302**, 716-718 (1983).
16. Neuberger, A., Gottschalk, A., Marshal, R. D. & Spiro, R. D. in *The Glycoproteins: Their*

Composition, Structure and Function Part A (ed. Gottschalk, A.) 450-490 (Elsevier, Amsterdam, 1972).

17. Proudfoot, N. J. & Brownlee, G. G. *Nature* **263**, 211-214 (1976).
18. Pelham, H. R. B. & Jackson, R. J. *Eur. J. Biochem.* **67**, 247-256 (1976).
19. Tien, M., Kirk, T. K., Bull, C. & Fee, J. A. *J. biol. Chem.* **261**, 1687-1693 (1986).
20. Kuila, D., Tien, M., Fee, J. A. & Ondrias, M. R. *Biochemistry* **24**, 3394-3397 (1986).
21. Poulos, T. L. *et al. J. biol. Chem.* **255**, 575-580 (1986).
22. Dunford, H. B. & Stillman, J. S. *Coord. Chem. Rev.* **19**, 187-251 (1976).
23. Poulos, T. L. & Kraut, J. *J. biol. Chem.* **255**, 8199-8205 (1980).
24. Chance, B. *Arch. Biochem. Biophys.* **41**, 416-424 (1952).
25. Manthey, J. A. & Hager, L. P. *J. biol. Chem.* **260**, 9654-9659 (1985).
26. Zhang, Y. Z., Zylstra, G. J., Olsen, R. H. & Reddy, C. A. *Biochem. biophys. Res. Commun.* **137**, 649-656 (1986).
27. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. *Biochemistry* **18**, 5294-5299 (1979).
28. Bantle, J. A., Maxwell, I. H. & Hahn, W. E. *Analyt. Biochem.* **72**, 413-427 (1976).
29. Tu, C.-P. D., Weiss, M. J., Karakawa, W. W. & Reddy, C. C. *Nucleic Acids Res.* **10**, 5407-5419 (1982).
30. Laemmli, U. K. *Nature* **227**, 680-685 (1970).
31. Towbin, H., Staehelin, T. & Gordon, J. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4350-4354 (1979).
32. Welinder, K. G. & Mazza, G. *Eur. J. Biochem.* **73**, 353-358 (1977).
33. Welinder, K. G. *FEBS Lett.* **72**, 19-23 (1976).

Corrigendum

A cDNA clone encoding an S-locus-specific glycoprotein from *Brassica oleracea*

J. B. Nasrallah, T.-H. Kao, M. L. Goldberg & M. E. Nasrallah
Nature **318**, 263-267 (1985).

FURTHER work on the sequence analysis of λ gt 10 cDNA clones related to pBOS5 has revealed that the nucleotide sequence of the cDNA clone should have:

- an insertion of a C after position 333;
- a change from CG to GC at positions 347-348;
- an insertion of G after position 766;
- an insertion of an A after position 808.

The reading frame is correct as published but the above insertions extend it to the TAG at position 1,252-1,254 raising the number of amino acid residues coded by this cDNA to 418.

MATTERS ARISING

Evidence for climatic attractors

GRASSBERGER¹ draws attention to some problems related to the estimate of attractor dimensions in the case of sparse data sets. As parts of this paper may appear to have a controversial character for the unaware reader, in reference to our previous work² we believe that some comments are in order.

It is surprising that with a number of points as small as 230 for the V28-238 core, Grassberger goes to phase spaces of dimension up to 11. In our own work with ~500 data points we reluctantly went up to dimension six, and considered that beyond this value the results were not reliable. In particular, the estimate of the slopes becomes increasingly difficult—in fact after seeing Grassberger's paper we repeated our computations by taking only one out of two or out of three data points and found no significant trend toward saturation for embedding phase space dimension up to six. We believe therefore that the conclusions drawn by Grassberger with so few data points are subject to caution. In addition it seems unlikely that the V28-239 core can be used for dimension estimates. There are very few raw data points, considering that they span a time interval twice as long as the V28-238

core. Moreover, the measurements beyond 10⁶ yr are less reliable (P. Pestiaux, personal communication).

Dr Grassberger minimizes the role of time lag τ . This may be theoretically correct, but in practice many researchers have pointed out the importance of the choice of τ in the calculation of attractor dimensions^{3,4}. The argument appealing to correlation time is of course legitimate, but in many cases this time may be hard to estimate. It is unfortunate that no specific references regarding these points are provided by the author.

Dr Grassberger also warns against spuriously small dimension estimates arising from smoothed data. It should be realized that in real world complex systems direct access to physical variables may be impossible. In geology in particular, the connection between data (which refer to depth along the core) and physical variables (like temperature or ice volume) is established by appealing to a model. Depending on the type of model (which describes, among others, the sedimentation mechanism) a resolution of for instance, 2,000 yr (our data) or 5,000 yr (Grassberger's data) is achieved. In either case, however, an interpolation is involved, since the raw data need not be equidistant. We regard this as a fact of

life. As all current work on palaeo-climatology is based on such data, we do not see why we should refrain from using them.

In summary, these issues show how careful one has to be in discussing the behaviour of real world complex systems. They also highlight the need for abundant, high quality geological data. We iterate our strong conviction that climatic attractors do exist, a conviction that is strengthened by independent dimension estimates reported recently by Fraedrich⁵ and Saltzman and co-workers⁶.

C. NICOLIS

*Institut d'Aéronomie Spatiale
de Belgique,
Avenue Circulaire, 3
1180 Bruxelles, Belgium*

G. NICOLIS

*Faculté des Sciences de
Université Libre de Bruxelles,
Campus Plaine, CP 226,
Boulevard du Triomphe,
1050 Bruxelles, Belgium*

1. Grassberger, P. *Nature* **323**, 609-612 (1986).
2. Nicolis, C. & Nicolis, G. *Nature* **311**, 529-532 (1984).
3. Wolf, A., Swift, J. B., Swinney, H. L. & Vastano, J. A. *Physica D16*, 285-317 (1985).
4. Malmaison, B., Atten, P., Bergé, P. & Dubois, M. *J. phys. Lett. Paris* **44**, L897-L902 (1983).
5. Fraedrich, K. *J. Atmos. Sci.* **43**, 419-432 (1986).
6. Saltzman, B. in *Irreversible Phenomena and Dynamical Systems Analysis in Geosciences* (Reidel, Dordrecht, 1987).

GRASSBERGER REPLIES—In the original paper of Nicolis¹ and in the above reply to my criticism², the authors claim to have used 500 'data points'. They question my results, as I have only used 230 points. This is highly misleading. Both of our analyses were based on the same set of 184 data points, interpolated and smoothened by Berger and Pestiaux. What was different, was the interpolation. I was unable to obtain the version used in ref. 1 either from its authors or from the authors of ref. 1. Thus, I used the version given to me by Berger as the most recent one. If the precise interpolation is as crucial as is claimed by Nicolis and Nicolis, they only test the interpolation procedure but not any climatic attractor encoded in the original data.

The authors criticize my use of a small time lag τ , combined with high embedding dimensions n . They are wrong. One should take as short a τ as possible, keeping $\tau \times n$ constant. In all cases where too small a τ has created problems, either n was too small or one of the following was not taken into account: (1) use maximum instead of Euclidean normalization in the correlation sum; (2) do not include in the sum pairs whose time separation is less than the correlation time; (3) do not take any difference from white noise as a sign of determinism; (4) avoid measuring noise (not intrinsic noise) of a very particular type: very sparse and short high noise bursts over a low-noise background. For example, the conclusions of ref. 3 are wrong due to neglect of point (2).

Finally, the density of raw data points in core V28–239 is about the same as in V28–238. Even if points older than 1 Myr are doubtful, the more recent part of V28–239 shows no hint of a strange attractor either.

In summary, I must stress that I have no *a priori* conviction whether a climatic attractor exists or not. The data show no indication for it. This may only mean that they are bad signals, as one might indeed conjecture by comparing different data sets visually. But taken at face value, they exclude an attractor dimension less than ~ 4 .

P. GRASSBERGER

*Fachbereich 8 (Physik),
University of Wuppertal,
D5600 Wuppertal, FRG*

1. Nicolis, C. & Nicolis, G. *Nature* **311**, 529–532 (1984).
2. Grassberger, P. *Nature* **323**, 609–612 (1986).
3. Fraedrich, K. *J. atmos. Sci.* **43**, 419–432 (1986).

Oceanic eddy transports and satellite altimetry

A RECENT letter to *Nature*¹ suggested a disarmingly simple relationship between the root mean square (r.m.s.) streamfunction ψ for oceanic eddy motions and the eddy diffusivity K for a passive tracer. The letter also suggested a means whereby

these quantities might be estimated from repeat track satellite altimetry. Here I point out that the Mid-Ocean Dynamics Experiment² (MODE) supplied data adequate to validate the relationship between ψ and K . The same data set also allows a test of some assumptions made¹ that allowed ψ to be estimated from satellite altimetry data. This test suggests that estimates of ψ and subsequently of K and meridional heat fluxes may be too high.

From theoretical arguments³ the relationship between the zonal eddy diffusivity K_x and r.m.s. streamfunction ψ for depth-averaged flows can be estimated as $K_x = C\psi$, where C is a numerical constant close to 0.4. The meridional diffusivity is suppressed somewhat and is written as $K_y = C\psi(1 + (\mu\beta)^2)^{-1}$, where $\beta = \beta u/\zeta^2$, β is the meridional derivative of f (the Coriolis parameter), u is a characteristic velocity and ζ the r.m.s. relative vorticity. The factor μ , estimated¹ at ~ 3 , depends on the shape of the eddy spectrum and the degree of velocity anisotropy.

MODE² provided a detailed description of the kinematical fields in a small area of the North Atlantic Ocean. Specifically, descriptions are available of the statistics of the streamfunction field for the depth-averaged flows⁴ and the depth-dependent flows⁵, the vorticity fields⁶ and the lagrangian velocity field⁷. The lagrangian statistics were acquired by tracking SOFAR⁸ (sound fixing and ranging) floats drifting at a depth of 1,500 m. That depth is close to the zero crossing of the first internal wave mode, and so the float statistics are representative of the depth-averaged flow field.

The r.m.s. particle speed at a depth of 1,500 m in MODE was reported⁷ as 4.0 cm s^{-1} and the vorticity variance at the same depth⁶ as $4.1 \times 10^{-12} \text{ s}^{-2}$. Thus the factor β is near 0.2. The value of $\mu \sim 3$ presumes that Rossby wave dynamics dominate and that the zonal velocity variance u^2 exceeds the meridional, v^2 . In MODE it was in fact observed⁷ that $u^2 < v^2$ so the parameter should be reduced somewhat. Immediately we see that dispersion in the MODE area should be relatively isotropic, compared with the estimates for the North Pacific¹, as $1 + (\mu\beta)^2 < 1.4$ compared with the value of 2 used by Holloway¹. Henceforth let us ignore the anisotropy and test the simpler relationship $K = C\psi$. Freeland and Gould⁴ fitted simple models of the longitudinal and transverse velocity covariance functions ($f(r)$ and $g(r)$ respectively) to the 1,500 m depth velocity data. Freeland and Gould found that the models $f(r) = u^2(1 + br) \exp(-br)$ and $g(r) = u^2(1 + br - b^2r^2) \exp(-br)$ fitted well, where $u^2 = 8 \text{ cm}^2 \text{ s}^{-2}$ is the velocity variance and $b^{-1} = 31 \text{ km}$ is a correlation length scale. These functions are related to each other by $g(r) = d(f(r))/dr$ implying horizontally non-divergent flow and

the existence of a covariance function for the streamfunction $C(r) = \psi^2(1 + br + b^2r^2/3) \exp(-br)$, where the variance $\psi^2 = 3u^2/b^2$. Thus we estimate K_x and $K_y = 0.4\psi = 6.1 \times 10^6 \text{ cm}^2 \text{ s}^{-1}$. Using a more complex model of the covariances $f(r)$ and $g(r)$, McWilliams⁵ produces an independent estimate of $0.4\psi = 7.9 \times 10^6 \text{ cm}^2 \text{ s}^{-1}$. The dispersion of SOFAR floats at the same depth yields direct and independent estimates of the zonal and meridional eddy diffusivities⁷, namely $(K_x, K_y) = (7.8, 7.1) \times 10^6 \text{ cm}^2 \text{ s}^{-1}$. Thus we find excellent agreement between observations of K and estimates from 0.4ψ .

Holloway used satellite altimetry estimates of the variance of sea surface elevation in the North Pacific to estimate ψ^2 at the surface. To estimate a depth-averaged value of ψ^2 it was assumed that of the total near surface ψ^2 , "half is intensified near the surface and half is distributed over the depth of the water column". We have no means of checking this assumption in the North Pacific. However, in MODE the height fluctuation variance due to the barotropic mode was estimated⁵ at 1.5 cm^2 , and the variance in near-surface dynamic height (relative to 1,500 m) at 46 cm^2 . If we now assume that the depth-dependent part of streamfunction behaves like $\psi_1\phi_1(z)$ where $\phi_1(z)$ is the shape of the first internal wave mode⁹, then the variance in streamfunction as a function of depth is $\psi^2 = \psi_1^2\phi_1^2(z) + \psi_0^2$, where ψ_0^2 is the variance due to the barotropic mode. For MODE conditions we find the ratio of depth-averaged ψ^2 to surface $\psi^2 \approx 0.18$ rather than the value¹ of 0.5. If that result is typical of the world ocean then the subsequent estimates of the diffusivity and eddy heat flux may be too high by a factor of 2 to 3. Holloway¹ notes the possibility that currents are more surface intensified than implied by his ratio of 0.5 but also that the discounting of upper ocean heat transport may be too severe.

HOWARD FREELAND

*Institute of Ocean Sciences,
Sidney, British Columbia,
Canada, V8L 4B2*

1. Holloway, G. *Nature* **323**, 243–244 (1986).
2. The MODE Group *Deep-Sea Res.* **25**, 859–910 (1978).
3. Holloway, G. & Kristmannson, S. S. *J. Fluid Mech.* **141**, 27–50 (1984).
4. Freeland, H. J. & Gould, W. J. *Deep-Sea Res.* **23**, 915–923 (1976).
5. McWilliams, J. C. *J. phys. Oceanogr.* **6**, 810–827 (1976).
6. McWilliams, J. C. *J. phys. Oceanogr.* **6**, 828–846 (1976).
7. Freeland, H. J., Rhines, P. B. & Rossby, J. T. *J. mar. Res.* **33**, 383–404 (1975).
8. Rossby, H. T., Voorhis, A. D. & Webb, D. *J. mar. Res.* **33**, 355–382 (1975).
9. Richman, J. *MODE Hot Line News* No. 12, 1–4 (1972).

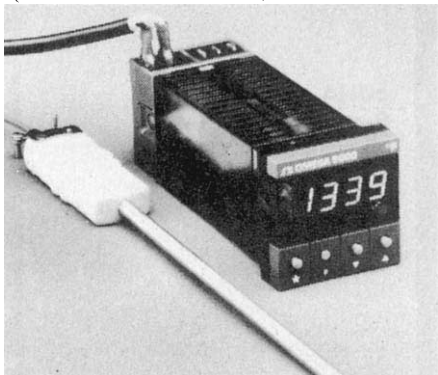
Matters Arising

Future contributions to Matters Arising should instead be sent to Scientific Correspondence and need not be sent to the authors of the original article in advance. Priority will be given to letters of less than 500 words and five references.

New ideas for the laboratory

This month's New of the Market spotlights new ideas for old problems, including paper to sop spills, a portable fume hood and a new perspective on microscopy.

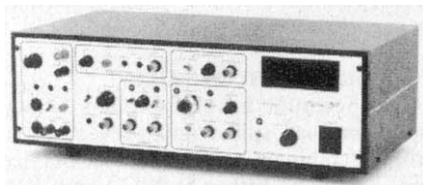
OMEGA Engineering has shrunk its **temperature controller** down so small that it practically fits in a pocket. Omega's CN9000 digital temperature controller employs a microprocessor to achieve its miniature size — only 1.88 inches high, 1.88 inches wide and 5.03 inches deep (Reader Service No. 101). The controller's



Omega's controller is miniaturized with a microprocessor.

LED display is large for its size at 0.4 inches, and is complemented by auxiliary indicators for outputs one and two, with three additional LEDs to indicate deviation from the setpoint. Each unit has dual outputs, with either a relay or voltage pulse output. And at \$195 (US), Omega's pocket-sized controller won't empty buyer's pockets.

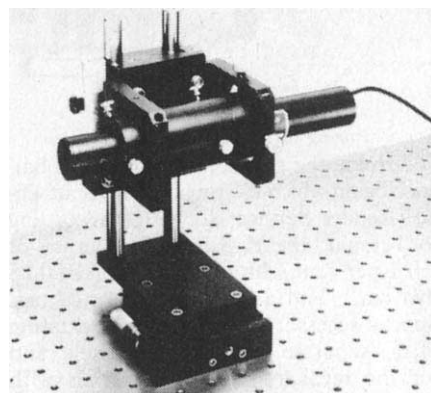
From World Precision Instruments, Inc. comes a new **intracellular amplifier**, the Intra 767 (Reader Service No. 102). The \$2,500 (US) low-noise intracellular electrometer features a miniature gold-



For use in neuroscience and nerve cell studies.

plated active microprobe preamplifier that weighs only 14 g and can be mounted on most micromanipulators. The Intra 767's LED meter makes the reading easy to see, and a differential amplifier can difference any applied signal with the 767's headstage output.

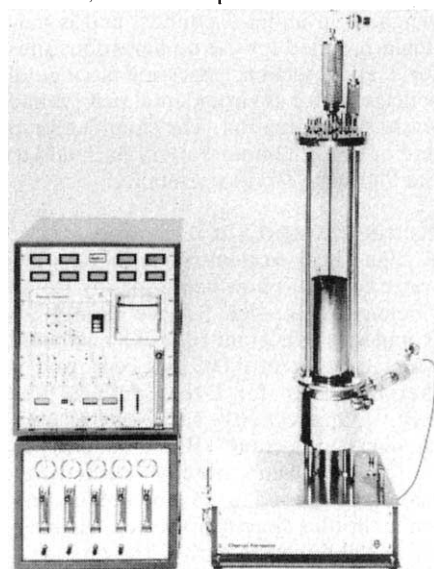
For pointing lasers in the right direction, Oriel Corporation offers a line of **precision laser pointers** (Reader Service No. 103). The \$855 (US) pointers fit all tubular laser diameters from 32 mm to 44 mm, and mount on tables or carriers with 6 mm on 50 mm centres. Angular adjustment is accomplished by mounting the



Oriel's pointer gets lasers pointed in the right direction.

front of the laser head in a gimbal mount, positioning the laser back with two precision thumbscrews for separate vertical/horizontal adjustment. The horizontal/vertical translators consist of pre-stressed ball bearings riding on hardened steel ways, and are driven by metric micrometers. Motorized translator drives may be purchased as options, and include Oriel's Motor, Encoder and Stepper Mike drives.

Alfa-Laval's **fermenters have now gone modular**. Its Compact Fermentation systems enable biotechnologists to build instruments, using a range of modular components, to meet the precise requirements of their experiment (Reader Service No. 104). The fermenter range includes *in situ* sterilizable airlift and stirred tanks from 3.5 to 38 litres, with a choice of seven interchangeable agitators. Agitation systems include the Effigis turbine, FUNDA-Spin stirrer and Vibro-



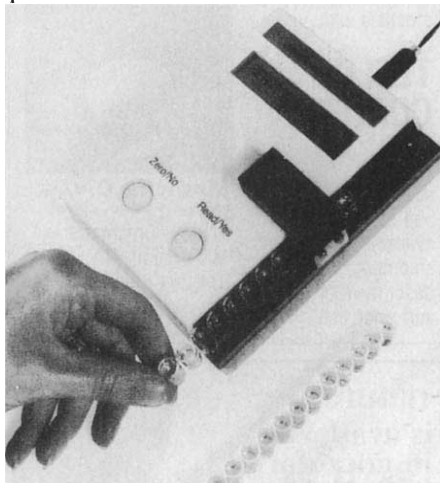
An example of a possible configuration using Alfa-Laval's fermenters.

mixer. Flexibility is afforded by the capability of Alfa-Laval's compact base units and vessels to be fitted to standard existing laboratory fermentation equipment.

Spectra-Physics has a new **Nd:YAG pulsed laser** that is a third the size of previous lasers, requires no cooling water, and has a small book-sized remote control module (Reader Service No. 105). The \$25,500 (US) DCR-11 is designed for portable applications, such as LIDAR in aircraft, and has an energy output of 275 mJ at 1,064 nm at a repetition rate of 10 p.p.s. The laser head measures $33 \times 9 \times 10$ inches, and the power supply is $23 \times 19 \times 21$ inches. The user may select 8 or 2.5 ns pulsewidths, and Spectra-Physics claims beam divergence and pointing stability are better than 0.5 mRad.

Convenient solutions

Whatman has made **colorimetry more convenient** with the introduction of its μ -Colorimeter (Reader Service No. 106). The £599 (UK) unit reads the absorbance or concentration of samples in a μ -cuvette strip, and the results can be recorded on the Whatman dot matrix printer or other printers that have an RS232 interface.



Whatman's idea to speed up colorimetry.

Operation of the μ -Colorimeter is simplified because the user has only two responses to the LCD readout: Read/Yes or Zero/No. Whatman says its machine covers an absorbance range of 0–2.000 Å with a 0.001 Å resolution and 2 per cent accuracy, and the concentration range is 0–999 and is supplied with a 405 nm filter and mains adaptor. Other filters, covering the range 400–700 nm, may be purchased as options.

Tekmar is offering a manually operated **surface and interfacial tensiometer**, the K-8, that uses a non-contacting sensor to

Water is the Basis of Life

Thales, 620 to 540

Seradest®

Water Deionisation by
Ion Exchange

- Disposable Systems
- Service Regeneration
- Self-Regeneration

Water Demineralisation by
Membrane Separation

- Reverse Osmosis

Seralpur®

Water Purification by Multistage
Post-Treatment

We are prepared for your
demands. Just ask us for the
most appropriate system for you.



Water Purification Systems
for Analysis, Production, Engineering

SERAL E. Alhuser GmbH, P.O. Box 551
D-5412 Rensbach-Baumbach/W.-Germany
Phone 0 26 23 - 8 91 - 0, Tlx. (17) 2623955
Telefax 02623-891-220

Service in

(A) (B) (CH) (DK) (F) (GB) (I) (N) (NL) (S) (SF)

Reader Service No.25

Medwire Div.

stocks a wide variety
of

TEFLON COATED*

Platinum alloys,
stainless steel,
silver, multistrand
and single strand
available for prompt
shipment. Write for
descriptive brochure
and price lists.



Medwire Div.

Affiliate of Sigmund Cohn Corp
123 So. Columbus Ave.
Mt. Vernon, N.Y. 10553
914-664-5300

*Reg. T.M. of E.I. DuPont

Reader Service No.49

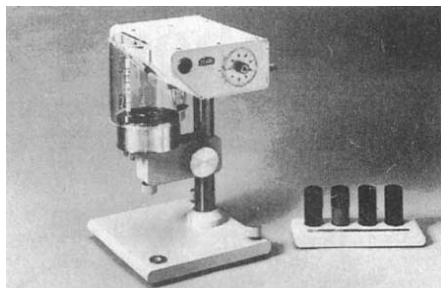
nature is available in microform from University Microfilms International.

Call toll-free 800-521-3044. In Michigan,
Alaska and Hawaii call collect 313-761-4700. Or
mail inquiry to: University Microfilms International,
300 North Zeeb Road, Ann Arbor, MI 48106.



Setting it straight

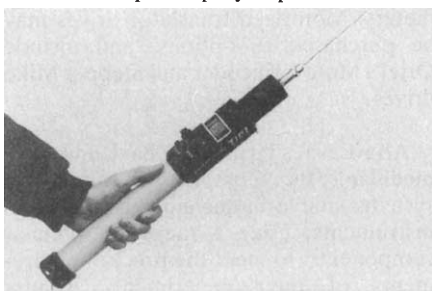
A REVIEW published in the 29 January issue of *Nature* (325, p. 467) stated that Technicol had recently developed a new line of columns to aid the flash chromatographer. The columns were in fact developed by May & Baker Ltd, and are distributed by Technicol.



For measuring surface and interfacial tension.

determine the position of a precision balance beam which the ring or plate is attached (Reader Service No. 107). According to Tekmar, the K-8 has an accuracy of ± 0.1 per cent with a resolution of 0.05 dyn cm^{-1} , and a \$5,000 (US) price tag. Special accessories to allow thermostating of the separate glass sample vessels, and for the measurement of materials with sample volumes of as small as 0.2 ml are available.

Even though it looks like a laser sabre from 'Star Wars', the Photovac Tip 1 from Centronic is really a **self-contained probe** for detecting air and water-borne contamination (Reader Service No. 108). The Photovac Tip 1 employs a photoionization



Probes the environment with a photoionization detector.

detector with a linear span range of between 0.1 and 2,000 p.p.m., and displays the results on an LCD. Centronic says the probe is designed for use in the field, weighing in under 3 pounds, and is particularly suited for use on hazardous sites for leak detection, assessing accidental spillage, or for environmental surveys and waste site evaluation. The shoulder strap and optional 32-hour battery pack add to the Photovac Tip 1's portability.

Reagents and such

A high-purity **acetonitrile with very low water content** is now being sold by Romil Chemicals (Reader Service No. 109). Romil says its acetonitrile has a maximum water content of 0.005 per cent, and is therefore ideal for DNA synthesis and HPLC applications such as electrochemical detection. Romil makes its acetonitrile from carefully selected raw materials, filtered to $0.5 \mu\text{m}$, and fills new amber bottles directly from the glass distillation still to ensure purity. The product is purged with pure nitrogen during bottling for prolonged shelf-life and to prevent

build-up of ultraviolet-absorbing impurities. Available in 2.5-litre bottles, Romil's acetonitrile costs £15.40 (UK).

From Bethesda Research Laboratories (BRL) comes a new **cDNA Synthesis System** (Reader Service No. 110). BRL claims researchers who use its **cDNA synthesis system** can retain more sequence information from the 5' end of the mRNA, and sequence as much as 40 per cent full-length double-stranded cDNA. The

cDNA Synthesis System

Instruction Manual

Cat No. B2675A

BRL

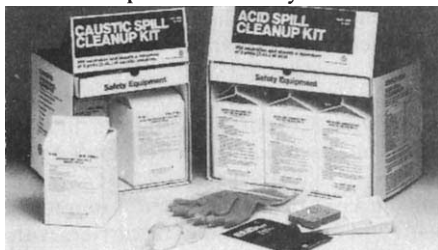
Bethesda Research Laboratories • Life Technologies, Inc.

The manual for BRL's cDNA synthesis system.

system is incorporated into one tube, and uses prequalified components, saving time and effort. The system's manual describes procedures for cloning into either plasmid or phage vectors. The cloning enzyme used is Moloney murine leukaemia virus reverse transcriptase, that BRL claims is free of detectable DNA endonuclease and contaminating RNase. Each system contains enough reagent for five operations, with 2–10 μg of mRNA per reaction.

Striving towards safety

Fisher Scientific has the solution to **mop up laboratory spills**, and prevent them from becoming a hazard (Reader Service No. 111). The Fisher Spill Cleanup Kits are outlined in a free product bulletin available from Fisher that describes the company's newest three emergency packs for handling flammable solvent, caustic, and acid spills. Each kit discussed in the brochure provides ready-to-use neu-



Fisher's kits come in convenient boxes to keep near at hand.

ralizer/absorbent material, disposal bags, scoops, safety goggles, gloves and other gear, and comes in a compact compartmented carton. For bigger accidents, the booklet also includes information on a bulk-sized version of the acid cleanup kit, that comes in a 50 kg drum.

Schleicher & Schuell lays the table for lab benches with its Polytrap 296 PE **bench-lining paper** (Reader Service No. 112). The absorbent polyethylene coated paper sops up spills, and protects lab surfaces from contamination. For working with pathogens and microorganisms, the



Really soaks up accidental laboratory spills. paper may be soaked with disinfectants. And Polytrap 296 PE makes valuable spills easier to recover, because it's simpler to recover spills from the absorbent paper than from the joints of a lab bench. Schleicher & Schuell's paper comes in rolls up to 50 m in length with widths of 40, 60, or 120 cm, or in sheets 48 × 60 cm in size.

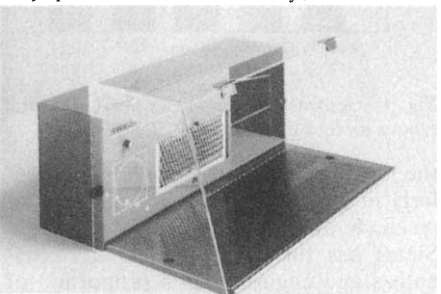
Transferring liquids safely and accurately need not present a problem, now that English Glass has developed the Englass Major (Reader Service No. 113). The new dispenser pump is designed for convenience and ergonomics. It fits on the outside of the container and draws the liquid up through a dip-tube, keeping the



The solution for convenient liquid transfer.

nozzle stationary during pumping. The Englass Major is available with different dispensing amounts, from 25 ml to 100 ml, in steps of 10 ml. Hazardous materials can be handled easily and safely because the risk of spillage and splashing is reduced, and the pump features a drip-free valve. The dispenser pump can be used for containers ranging in size from 2 litres to 250 litres, and Englass will print brands or instructional copy on the external fitting of the pump for large orders.

Shandon's Hyperclean-1 **fume hood** allows researchers to take their work with them (Reader Service No. 114). The Hyperclean-1 is a portable ductless system that can be conveniently moved to any part of the laboratory, as needed.

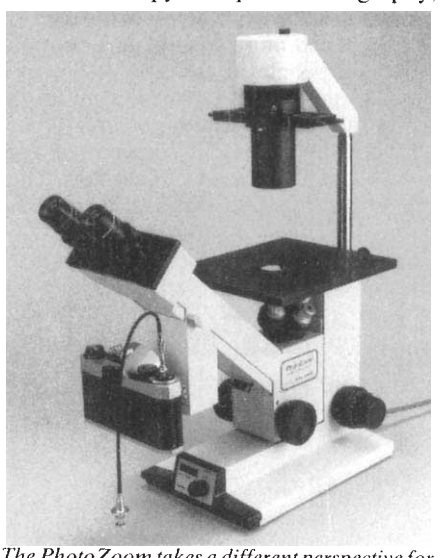


Shandon's fume hood's design makes it easily portable.

Fumes are absorbed by a charcoal filter cartridge that can be replaced when spent, and the hood has a corrosion-resistant fan. The £780 (UK) hood is transparent, to take advantage of existing light, and has two removable acrylic shelves. A spill tray lines the bottom of the hood, and can be removed for cleaning.

Imaging ideas

Bausch & Lomb has turned microscopy upside down, and come up with its PhotoZoom **inverted microscope** (Reader Service No. 115). Specially designed to meet the complex requirements of tissue culture microscopy and photomicrography,



The PhotoZoom takes a different perspective for microscopy.

the \$2,279–\$2981 (US) PhotoZoom has a built-in parfocal × 0.8–× 2.9 zoom system for low magnification scanning with a 25 mm field of view. A fluid-proof nosepiece protects the microscope's internal workings from spills, and an integral port for photomicrography adapts to all T-Mount 35mm camera systems and C-Mount video cameras. The PhotoZoom is distributed exclusively in the United States by American Scientific Products, and by authorized dealers in the international marketplace.

A new film by Polaroid for **radiography and non-destructive testing** can be developed in 45 seconds without a dark-room or chemicals (Reader Service No. 116). The 8 × 10 inch black-and-white Type 803 film can be exposed in portable,



Fast-developing radiography film from Polaroid.

cabinet or real-time X-ray units, and then processed in normal room light with the portable Polaroid 8 × 10 film processor. Prints require no coating after processing to protect them and ensure long-term image stability. The Type 803 film costs \$139.75 (US) for 15 exposures. □

Coming in Product Review...

- **Labelling techniques.** The 16 April issue will contain a review of monoclonal antibodies, fluorescent probes, and other labelling reagents, plus the equipment to detect them.
- **Laboratory software.** As a preview to the Scientific Computing and Automation Conference to be held in Amsterdam a week later, the 7 May issue will have a sampling of what's newest in software for science.
- **Tissue culture.** The 21 May Product Review will feature products to be displayed at the US Tissue Culture Association Meeting, 27–30 May, in Washington, D.C.

These notes are compiled by Carol Ezzell from information provided by the manufacturers. To obtain further details about these products, use the reader service card bound inside the journal. Prices quoted are sometimes nominal and apply only within the country indicated.

Research investment, pounds and people

Richard Pearson

Manpower shortage in research may cause problems for future investment.

INVESTMENT in research and development is seen as one of the key building blocks in developing a successful economy. This is despite uncertainty as to where such resources can be best directed and how research can best be linked to development and to the market place. All these issues are being hotly debated worldwide, with Britain no exception. Some dozen major reports addressed these topics in the past year and the government established an Office of Science and Technology Assessment in the Cabinet Office and several new 'science policy' study units sprang up in other institutions. As a result of growing interest from researchers, lobbyists and politicians, an *Annual Review of Government Funded Research and Development* is now published each year, although it only focuses on government-funded research. In contrast the US National Science Foundation not only prepares reviews of the United States but has also begun a series of international reviews¹, the first on West Germany².

International comparisons can be made to prove many points. Some argue that the United Kingdom's expenditure, at around 2.2 per cent of the gross national product, is on a par with France and only slightly behind that of West Germany, Japan and the United States at 2.5 per cent. Others, however, say this is misleading because a far higher share, one-third, of the United Kingdom's expenditure is on defence which has little direct payoff for civilian activities. But in the United States where one-third of research and development is also related to defence, the effect is said to be beneficial with a strong positive link between defence and civil applications.

Employment statistics provide another basis for comparisons and a recent five-country review showed that Britain has the highest density of scientists and engineers at 25 per thousand of the working population and also the highest participation rate by women (mainly scientific), with the United States coming second in each case (Fig. 1). Among these five countries Japan has the highest proportion of young scientists and engineers, no doubt reflecting its late industrialization, although interestingly the United Kingdom ranks second with almost half aged under 35, while in the other countries the proportions are nearer one in three. These aggregate figures show Britain in a favourable light, but obscure the high density of scientists in manufacturing and

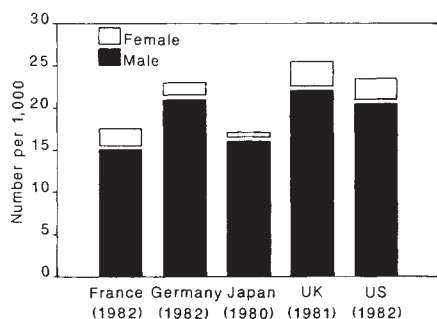


Fig. 1 Scientists and engineers per 1,000 total labour force, by sex (from ref. 2).

the low proportion of engineers, particularly in electronics (Fig. 2). In terms of research and development the United States has the greatest number of scientists and engineers as a proportion of the labour force with 65 per thousand but Japan is catching up fast. The United Kingdom with 35 per thousand suffers badly by comparison.

These statistics focus on the early 1980s, but what of the future? This will be determined both by expenditure on research and science and technology, and the availability of suitably trained and skilled workers. The link between investment in education and economic success, while not proved, is generally accepted and most of the developed and new industrialized nations are now heavily investing in education as well as research and development to achieve economic goals. In the United Kingdom, however, there is strength in the output of graduate scientists but relative weakness in the case of engineers and more general participation by young people in education³. For example 69 per cent of Japanese aged 16–18 are in full-time education, as are 79 per cent of Americans, 58 per cent of the French and 45 per cent of Germans. The

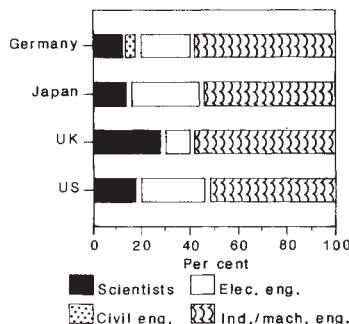


Fig. 2 Distribution of scientists and engineers in manufacturing, by occupation group.

figure for the United Kingdom is only 32 per cent although there is a relatively high proportion, 32 per cent, in part-time education or training. Figures for higher education show a participation rate of 31 per cent, if part-timers are included, which is roughly on a par with both our European competitors and Japan but is only half that of the United States.

Whatever the level of investment in education the key resource is, of course, people and perhaps the biggest challenge of the 1990s is for Europe and North America to cope with the forthcoming huge decline in college-age manpower supply, in contrast to Japan where it is increasing. For example between 1980 and 1994 the number of 16–18-year-olds in the population will fall by over 20 per cent in the United States while here the fall will

Table 1 Index of 16–18-year-old population 1980–1994 (Index 1980 = 100)

Year	France	FRG	Japan	UK	USA
1980	100	100	100	100	100
1982	102	102	101	102	92
1984	99	100	105	98	87
1986	98	93	115	95	88
1988	99	80	122	91	83
1990	100	66	125	82	76
1992	92	58	119	74	76
1994	86	56	109	74	77

Source: ref. 1.

be 36 per cent. The most extreme fall however is occurring in West Germany where the fall will be nearer 45 per cent (Table 1).

In the United States the supply of scientists and technologists may be further weakened by the reported swing in student preferences towards business and the professions and away from science and technology. The demographic downturn can, in part, be attributed to rising education levels by women and improved contraception as a result of technological advances. It would be ironic indeed if the two factors seen as the key to prosperity in the next century (investment in research and development and in education) could be the very factors which will induce skill shortages in the run-up to the year 2000. □

Richard Pearson is at the Institute of Manpower Studies, Mantell Building, University of Sussex, Falmer, Brighton BN1 9RF, UK

1. *International Science and Technology Data Update* (National Science Foundation, 1986).
2. *Science and Technology Resources of West Germany* (National Science Foundation, 1986).
3. Pearson, R. *Nature* 313, 82 (1985).